

High-Rate Vector Quantization for the Neyman-Pearson Detection of Correlated Processes

Joffrey Villard, *Student Member, IEEE*, and Pascal Bianchi, *Member, IEEE*

Abstract

This paper investigates the effect of quantization on the performance of the Neyman-Pearson test. It is assumed that a sensing unit observes samples of a correlated stationary ergodic multivariate process. Each sample is passed through an N -point quantizer and transmitted to a decision device which performs a binary hypothesis test. For any false alarm level, it is shown that the miss probability of the Neyman-Pearson test converges to zero exponentially as the number of samples tends to infinity, assuming that the observed process satisfies certain mixing conditions. The main contribution of this paper is to provide a compact closed-form expression of the error exponent in the high-rate regime *i.e.*, when the number N of quantization levels tends to infinity, generalizing previous results of Gupta and Hero to the case of non-independent observations. If d represents the dimension of one sample, it is proved that the error exponent converges at rate $N^{2/d}$ to the one obtained in the absence of quantization. As an application, relevant high-rate quantization strategies which lead to a large error exponent are determined. Numerical results indicate that the proposed quantization rule can yield better performance than existing ones in terms of detection error.

Index Terms

Binary hypothesis testing, compression, error exponents, hidden Markov models, stochastic processes, vector quantization.

J. Villard is with the Department of Telecommunications, SUPELEC, 91192 Gif-sur-Yvette, France (e-mail: joffrey.villard@supelec.fr).

P. Bianchi is with Telecom ParisTech, 75634 Paris Cedex 13, France (e-mail: bianchi@telecom-paristech.fr).

The work of J. Villard is supported by DGA (French Armement Procurement Agency).

I. INTRODUCTION

Consider a sensing unit which transmits a sequence of measurements to a decision device (DD) whose mission is to detect a given signal. For example, a CCTV camera in a surveillance system transmits its data to a remote controller interested in the detection of a particular object in its field of view. This situation also arises in the context of wireless sensor networks (WSN) where a fusion center collects the individual measurements of a large number of identical sensors and processes these measurements in order to detect abnormal events [1], [2]. In such applications, due to bandwidth, delay or storage limitations, transmitted data rates are often limited. Therefore, measurements must be quantized prior to transmission. As a matter of fact, this quantization step may severely degrade the overall detection performance of the system.

In this paper, we consider that a binary hypothesis test is performed at the DD. The available data set corresponds to a quantized version of a stationary ergodic discrete-time multivariate process. Our aim is to quantify the detection performance of a given quantizer and characterize quantization strategies which guarantee attractive performance at the DD.

In the past decades, numerous papers were dedicated to the search for relevant quantization strategies and their practical design [3]. The most popular criterion used to select quantizers is the mean square error (MSE) between the quantized signal and the initial source [4]. An analytical characterization of quantizers minimizing the MSE is difficult in the general case. Bennett [5] pioneered the study of *high-rate* (or *high-resolution*) quantization for the reconstruction of scalar signals. The idea of Bennett was to study the MSE in the asymptotic regime where the number of quantization levels tends to infinity. A closed form expression of the (properly normalized) MSE can be determined in that case, and the families of quantizers minimizing the asymptotic MSE can be directly characterized. Extension of the work of Bennett to vector-valued observations was later achieved in [6]. However, the MSE criterion is especially relevant when the aim is to reconstruct the source. On the other hand, it can be inappropriate as far as other applications are concerned. For this reason, various distortion measures have been proposed in the literature in a *task-oriented* setting for estimation, classification and detection [7]–[18]. In particular, considerable attention has been paid to optimal quantization for hypothesis testing. Poor and Thomas [12] used Ali-Silvey distances between densities. Later, Poor [13] proposed the generalized f -divergence and studied this distortion measure in the high-rate regime. Picinbono

and Duvaut [14] considered a deflection criterion and proved that the corresponding optimal procedure corresponds to the scalar quantization of the likelihood ratio. Tsitsiklis [15] studied the properties of such quantizers with respect to several distortion measures. More recently, following the initial works of Tenney and Sandell [16] and Tsitsiklis [17], Gupta and Hero [18] investigated the selection of high-rate quantizers for binary hypothesis tests. In their setting, the decision device gathers a sequence of n independent and identically distributed (i.i.d.) variables, each of these variables being passed through a fixed quantizer. The probability density function (pdf) of the samples is assumed to be known both under the null hypothesis and the alternative. In this case, it is well known that a uniformly most powerful test is obtained by the Neyman-Pearson (NP) procedure which consists in rejecting the null hypothesis when the log-likelihood ratio (LLR) exceeds a certain threshold [19]. The threshold is usually chosen in such a way that the probability of false alarm of the test (that is, the probability to decide the alternative under the null hypothesis) is fixed to a specified *level*, say α . The performance of the NP test of level α can be evaluated in terms of the miss probability (that is, the probability to decide the null hypothesis under the alternative). In our case, the miss probability clearly depends on the quantizer used by the sensing unit. Thus, a natural approach would be to select the quantizer which minimizes the miss probability. Unfortunately, the miss probability does not admit any tractable expression as a function of the quantizer. To circumvent this issue, it is convenient to study the miss probability in the case where the number n of available snapshots tends to infinity. In case of i.i.d. observations, the celebrated Stein's lemma [20] states that the miss probability tends to zero exponentially in n . Based on this result, it is relevant to select the quantizers which yield a large value of the error exponent. Unfortunately, the maximization of the error exponent as a function of the quantizer is impractical. Following the idea of [5], [6], Gupta and Hero restrict their attention to high-rate quantizers and manage to obtain a compact expression of the error exponent loss induced by quantization.

Most of these works address the case where observations are independent random variables. However, the detection of a correlated process is a crucial issue in many applications [21]–[24]. In this case, fewer results are available in the literature. Chamberland and Veeravalli [21] analyze the impact of the density of sensors in a WSN on the detection performance, when observations are correlated. Willett *et al.* [22] study the one-bit quantization of a pair of dependent Gaussian random variables. In case of the detection of a Gauss-Markov signal in noise, Sung *et al.* [23]

prove that for a fixed false alarm level, the miss probability of the NP test converges exponentially to zero, and provide a closed form expression of the error exponent. Hachem *et al.* [24] later extended the results of [23] to irregularly sampled Gaussian diffusion processes. However, [23], [24] assume that the DD has a perfect access to the observations of the sensing unit, and do not address quantization issues.

In this paper, we study the performance of the Neyman-Pearson test based on a quantized version of a stationary ergodic multivariate process. We generalize the work of Gupta and Hero [18] to the case where the observed process is non-i.i.d. (either under the null hypothesis, the alternative, or both). In this situation, Stein's lemma does not directly apply. The error exponent does no longer admit a closed-form expression and the determination of relevant quantizers is therefore a more difficult task. Provided that the process of interest satisfies certain forgetting properties (present observations should become nearly independent of past observations after a sufficient amount of time), we prove that the miss probability of the NP test of level α tends exponentially to zero as the number of observations tends to infinity. Our main contribution is to provide a compact closed form expression of the error exponent in case of high-rate quantizers. If N denotes the number of quantization levels (or equivalently if each measurement is quantized on $\log_2(N)$ bits), we prove that the error exponent achieved when using quantized observations converges as N tends to infinity to the ideal error exponent that one would obtain if perfect/unquantized measurements were available at the DD. More precisely, we prove that the error exponent loss tends to zero at speed $N^{-2/d}$ where d represents the dimension of each individual measurement. The asymptotic error exponent depends on the process distributions under both hypotheses. It also depends on the quantization strategy through the so-called *model point density* and *model covariation profile*. The model point density can be interpreted as the asymptotic density of cells in the neighborhood of each point of the observation space. The model covariation profile captures the shape of the cells. As a consequence, the selection of relevant high-rate quantizers reduces to the determination of the point densities and covariation profiles minimizing the asymptotic error exponent loss. In case of scalar quantization ($d = 1$), our compact expression immediately yields a simple characterization of optimal high-rate quantizers. In case of vector quantization ($d \geq 2$), an exact characterization of optimal quantizers is more difficult. Following the approach of [18] once again, we nevertheless determine relevant families of quantizers with attractive error exponent. Note that our theoretical results

hold under the assumption that the observed process “forgets” past observations fast enough. As a special case, we prove that our assumptions hold for a general class of hidden Markov models verifying a certain contraction property. Numerical illustrations are provided in the case where the measurements correspond to a modulated signal in the In-phase/Quadrature plane.

The paper is organized as follows. In Section II, we describe the observation model. We also review some known results on Neyman-Pearson tests and we derive the associated error exponent in the ideal case where the DD has perfect access to the measurements. The vector quantization framework is introduced in Section III. In Section IV, the impact of quantization on the error exponent is evaluated in the high-rate regime. We determine relevant quantization strategies allowing to reduce this degradation. Section V is devoted to the proof of the main result. In Section VI, we illustrate our findings in the special case of hidden Markov processes and give sufficient conditions on the transition and observation kernels ensuring that our results apply. Section VII is dedicated to numerical illustrations.

Notation

For any sequence $(y_i)_{i \in \mathbb{Z}}$, for any integers $k \leq \ell$, notation $y_{k:\ell}$ stands for the collection $(y_k, y_{k+1}, \dots, y_\ell)$ and notation $y_{\mathbb{Z}}$ is used to designate the whole sequence. If y is a vector with dimension d , we denote by $y^{(i)}$ its i -th component and $\|y\|$ its Euclidean norm. We denote by $\|A\|$ the spectral norm of any square matrix A . Notation T stands for the transpose operator.

A real-valued function $f : y_{k:\ell} \mapsto f(y_{k:\ell})$ on $S \subset \mathbb{R}^d \times \dots \times \mathbb{R}^d$ is said to be of class C_3 on S if it is three times continuously differentiable on S . We denote by $\nabla_{y_m} f(y_{k:\ell})$ its gradient w.r.t. y_m at point $y_{k:\ell}$. When no variable is specified, $\nabla g(y)$ simply denotes the (d -dimensional) gradient of the real-valued single-variable function $y \mapsto g(y)$ defined on $Y \subset \mathbb{R}^d$. We define the Hessian matrix of f by $[\nabla_{y_m, y_n}^2 f]_{i,j} = \frac{\partial^2 f}{\partial y_m^{(i)} \partial y_n^{(j)}}$ for all $i, j \in \{1, \dots, d\}$. Moreover, notation $\nabla_{y_m}^2$ stands for ∇_{y_m, y_m}^2 .

Notation $B(X)$ stands for the Borel σ -field on X . Notation $\sigma(Y_{1:n})$ stands for the sub- σ -field of $B(Y_{\mathbb{Z}})$, associated with the random vector $Y_{1:n}$. Notation $\xrightarrow[n \rightarrow \infty]{P}$ stands for the convergence in probability as $n \rightarrow \infty$. Notation $\xrightarrow[n \rightarrow \infty]{L^r(\mathbb{P}_0)}$ stands for the convergence in the L^r -norm w.r.t. probability \mathbb{P}_0 .

Notation \circ stands for the composition operator *i.e.*, for any arbitrary functions f and g , $f \circ g(x) = f(g(x))$. Notation $o_N(\cdot)$ is a little-o notation as N tends to infinity.

II. NEYMAN-PEARSON DETECTION WITH PERFECT OBSERVATIONS

A. Observation Model

Consider two probability measures \mathbb{P}_0 and \mathbb{P}_1 on a relevant probability space. Denote by $(Y_k)_{k \in \mathbb{Z}}$ a stationary ergodic process for both \mathbb{P}_0 and \mathbb{P}_1 , taking its values in a bounded convex subset Y of \mathbb{R}^d . We associate an hypothesis (H0 and H1 respectively) to each of the two probability measures \mathbb{P}_0 and \mathbb{P}_1 and investigate the problem of the detection of H1 vs. H0 based on a set of n observations $Y_{1:n} = (Y_1, \dots, Y_n)$.

For each $i \in \{0, 1\}$, we assume that \mathbb{P}_i is the probability distribution of the coordinate process $(Y_k)_{k \in \mathbb{Z}}$ on the canonical space $(Y^{\mathbb{Z}}, B(Y^{\mathbb{Z}}))$. We denote by $P_{i,n}$ the restriction of \mathbb{P}_i to $\sigma(Y_{1:n})$. We denote by \mathbb{E}_0 and \mathbb{E}_1 the expectations associated with \mathbb{P}_0 and \mathbb{P}_1 respectively. We introduce the reference measure μ which coincides with the d -dimensional Lebesgue measure restricted to Y .

Assumption 1: The following properties hold true for each $i \in \{0, 1\}$.

- 1) For each $n \geq 1$, $P_{i,n}$ admits a density p_i w.r.t. $\mu^{\otimes n}$.
- 2) $p_i(y_{1:n}) > 0$ for each $y_{1:n} \in Y^n$.
- 3) $\mathbb{E}_0 |\log p_i(Y_0)| < \infty$.

The density p_i of $P_{i,n}$ depends of course on n , but we drop the index n to simplify the notation. For each $i \in \{0, 1\}$, we also define $p_i(y_n | y_{1:n-1}) = p_i(y_{1:n}) / p_i(y_{1:n-1})$ with the convention that $p_i(y_n | y_{1:n-1}) = p_i(y_n)$ when $n = 1$ (that is, when $y_{1:n-1}$ is a void vector). Assumption 1-2) implies that both distributions $P_{0,n}$ and $P_{1,n}$ are absolutely continuous w.r.t. each other.

B. Likelihood Ratio Test

We now investigate the detection of H1 vs. H0 based on the perfect observation of n measurements $Y_{1:n}$. The log-likelihood ratio (LLR) writes:

$$L_n = \log \frac{p_1(Y_{1:n})}{p_0(Y_{1:n})}. \quad (1)$$

The NP test rejects the null hypothesis when L_n is larger than a threshold, say γ . For each $\alpha \in (0, 1)$, we define the miss probability of the NP test of level α by:

$$\beta_n(\alpha) = \inf \mathbb{P}_1 [L_n < \gamma] ,$$

where the infimum is w.r.t. all γ such that the probability of false alarm does not exceed α i.e.,

$$\gamma \text{ s.t. } \mathbb{P}_0 [L_n > \gamma] \leq \alpha .$$

For each $n \geq 1$ and each $\alpha \in (0, 1)$, due to the celebrated Neyman-Pearson's lemma, $\beta_n(\alpha)$ is the lowest achievable miss probability among all binary tests of level α which are based on the observation of $Y_{1:n}$. Quantity $\beta_n(\alpha)$ is therefore a key metric in order to characterize the performance of the hypothesis test. Unfortunately, it usually does not admit any tractable closed-form expression. In the sequel, we study the asymptotic behaviour of $\beta_n(\alpha)$ as the number of observations n tends to infinity. In this regime, it can be shown that, under certain assumptions,

$$\beta_n(\alpha) \simeq \exp(-n K) \quad (2)$$

for some constant K given below, which we shall refer to as the error exponent.

C. Error Exponent with Perfect Observations

The evaluation of the error exponent K in Equation (2) fundamentally relies on the following lemma:

Lemma 1 ([25]): Assume that a binary test is performed on a sequence $\check{Y}_{1:n} = (\check{Y}_1, \dots, \check{Y}_n)$ of n observed random variables. Denote by \check{p}_0 and \check{p}_1 the density of $\check{Y}_{1:n}$ under H_0 and H_1 respectively (w.r.t. any common reference measure). Assume that under H_0 ,

$$\frac{1}{n} \log \frac{\check{p}_0(\check{Y}_{1:n})}{\check{p}_1(\check{Y}_{1:n})} \xrightarrow[n \rightarrow \infty]{P} \kappa$$

for some deterministic constant κ such that $0 < \kappa \leq \infty$. Then, for any $\alpha \in (0, 1)$ the miss probability $\beta_n(\alpha)$ of the Neyman-Pearson test of level α is such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(\alpha) = -\kappa .$$

Lemma 1 implies that the error exponent, if it exists, coincides with the limit in probability (under \mathbb{P}_0) of $-(1/n)L_n$, where L_n is the LLR defined by (1). The existence of the error exponent is directly obtained from the following assumption, which will be discussed later on.

Assumption 2: For each $i \in \{0, 1\}$, $(\log p_i(Y_0|Y_{-m:-1}))_{m \geq 0}$ is a convergent sequence in $L^1(\mathbb{P}_0)$.

We are now in position to study the limit of the LLR L_n and prove the following result, which provides the general form of the error exponent.

Theorem 1: Under Assumptions 1 and 2,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(\alpha) = -K ,$$

where K is the constant defined by

$$K = \lim_{m \rightarrow \infty} \mathbb{E}_0 \left[\log \frac{p_0}{p_1}(Y_0|Y_{-m:-1}) \right] . \quad (3)$$

Proof: Using the chain rule, we first write L_n under the form:

$$L_n = - \sum_{k=1}^n \log \frac{p_0}{p_1}(Y_k|Y_{1:k-1}) .$$

Denote by Υ the limit in $L^1(\mathbb{P}_0)$ of sequence $(\log \frac{p_0}{p_1}(Y_0|Y_{-m:-1}))_{m \geq 0}$. The main point is the study of the difference $\log \frac{p_0}{p_1}(Y_k|Y_{1:k-1}) - \Upsilon \circ \theta^k$, where θ is the shift operator¹. We can write:

$$\begin{aligned} \mathbb{E}_0 \left| \frac{1}{n} L_n + \frac{1}{n} \sum_{k=1}^n \Upsilon \circ \theta^k \right| &\stackrel{(a)}{\leq} \frac{1}{n} \sum_{k=1}^n \mathbb{E}_0 \left| \log \frac{p_0}{p_1}(Y_k|Y_{1:k-1}) - \Upsilon \circ \theta^k \right| \\ &\stackrel{(b)}{\leq} \frac{1}{n} \sum_{k=1}^n \mathbb{E}_0 \left| \log \frac{p_0}{p_1}(Y_0|Y_{-k+1:-1}) - \Upsilon \right| , \end{aligned}$$

where step (a) comes from the triangular inequality and step (b) is a consequence of the stationarity of process $(Y_k)_{k \in \mathbb{Z}}$ under \mathbb{P}_0 . The right-hand side of the above inequality can be interpreted as a Cesàro mean and thus converges to zero by definition of Υ . We thus write:

$$-\frac{1}{n} L_n = \frac{1}{n} \sum_{k=1}^n \Upsilon \circ \theta^k + \varepsilon_n ,$$

where ε_n represents a term which converges in probability (under \mathbb{P}_0) to zero as $n \rightarrow \infty$. As \mathbb{P}_0 is stationary ergodic, we conclude using the ergodic theorem that $-(1/n)L_n$ converges in probability to $\mathbb{E}_0(\Upsilon)$ under \mathbb{P}_0 . This result together with Lemma 1 proves Theorem 1. ■

Remark 1: Let us make some remarks on the above Assumptions 1 and 2. Assumption 1 is an extension of those made by Gupta and Hero [18, Section III, pp. 1956]. Assumption 2 does not appear in [18] since it is obviously verified by i.i.d. processes. In this case, Theorem 1 is known

¹Recall that we are considering probability measures defined on the canonical space $\mathcal{Y}^{\mathbb{Z}}$. For any $\omega \in \mathcal{Y}^{\mathbb{Z}}$, we may write $\omega = (\dots, \omega_{-1}, \omega_0, \omega_1, \dots)$. The k th-time shifted version of ω is then given by $\theta^k \omega = (\dots, \omega_{k-1}, \omega_k, \omega_{k+1}, \dots)$. Notation $\Upsilon \circ \theta^k$ represents the measurable function $\Upsilon \circ \theta^k(\omega) = \Upsilon(\theta^k \omega) = \Upsilon((\dots, \omega_{k-1}, \omega_k, \omega_{k+1}, \dots))$. Recall that process $Y_{\mathbb{Z}}$ is defined as the coordinate process i.e., $Y_n(\omega) = \omega_n$ for each n . As a consequence, the measurable function $\log \frac{p_0}{p_1}(Y_k|Y_{1:k-1}) - \Upsilon \circ \theta^k$ at point ω is equal to the measurable function $\log \frac{p_0}{p_1}(Y_0|Y_{-k+1:-1}) - \Upsilon$ at point $\theta^k \omega$.

as Stein's lemma. Assumption 2 is trivially satisfied by short-dependent (m -dependent) processes such as moving average processes for instance [26]. In this case, the present observation Y_0 is independent of past observations $Y_{-m-1}, Y_{-m-2}, \dots$ as soon as m is large enough. As explained in Section VI, Assumption 2 is as well satisfied by a wide class of hidden Markov models.

Remark 2: In order that $(\log p_0(Y_0|Y_{-m:-1}))_{m \geq 0}$ is a convergent sequence in $L^1(\mathbb{P}_0)$, it is sufficient to check that $(\mathbb{E}_0 \log p_0(Y_0|Y_{-m:-1}))_{m \geq 0}$ is a bounded sequence. This claim is a consequence of Moy [27] (see Theorem 4 therein). In practical situations, this remark provides us with a convenient way to check whether Assumption 2 is verified for $i = 0$. On the other hand, the validation of Assumption 2 for $i = 1$ generally requires more efforts in practice: One should be able to prove that $(\log p_1(Y_0|Y_{-m:-1}))_{m \geq 0}$ is a Cauchy sequence in $L^1(\mathbb{P}_0)$.

Remark 3: When \mathbb{P}_1 is a finite-order Markovian measure, Assumption 2 can be simply reduced to the assumption that sequence $(\mathbb{E}_0 \log \frac{p_0}{p_1}(Y_0|Y_{-m:-1}))_{m \geq 0}$ is bounded. Indeed, due to Moy [27], this hypothesis directly implies the convergence of sequence $(\log \frac{p_0}{p_1}(Y_0|Y_{-m:-1}))_{m \geq 0}$ in $L^1(\mathbb{P}_0)$ and thus yields Theorem 1.

III. QUANTIZATION

A. Definitions

Consider a fixed integer $N \geq 2$. An N -point quantizer is a triplet $(\mathcal{C}_N, \Xi_N, \xi_N)$ where $\mathcal{C}_N = \{C_{N,1}, \dots, C_{N,N}\}$ is a set of N cells (Borel sets of \mathcal{Y} with non-zero volume) which form a partition of \mathcal{Y} , where $\Xi_N = \{\xi_{N,1}, \dots, \xi_{N,N}\}$ is an arbitrary set of distinct elements and where $\xi_N : \mathcal{Y} \rightarrow \Xi_N$ is a function s.t. $\xi_N(y) = \xi_{N,j}$ whenever $y \in C_{N,j}$. For each N, k , we introduce

$$Z_{N,k} = \xi_N(Y_k) ,$$

the quantized measurement on $\log_2(N)$ bits. We assume that the quantizer $(\mathcal{C}_N, \Xi_N, \xi_N)$ is known at the decision device. The aim is to decide between hypotheses H_0 and H_1 based on the observation of $Z_{N,1:n}$.

Note that in our model, each individual measurement is quantized based on the same quantization rule as in the traditional framework of vector-quantization [3]. It is also relevant in the case of WSN when samples are collected using identical sensors.

B. Error Exponent

Assume that the number of quantization levels N is fixed. For a given number n of quantized observations, we define the LLR based on quantized measurements by:

$$L_{n,N} = \log \frac{p_{1,N}(Z_{N,1:n})}{p_{0,N}(Z_{N,1:n})},$$

where for each $i \in \{0, 1\}$ and for any set of quantization points $\xi_{N,j1:n} = (\xi_{N,j1}, \dots, \xi_{N,jn}) \in \Xi_N^n$,

$$p_{i,N}(\xi_{N,j1:n}) = P_{i,n}(C_{N,j1} \times \dots \times C_{N,jn})$$

is the probability that measurements Y_1, \dots, Y_n respectively fall into the cells $C_{N,j1}, \dots, C_{N,jn}$ associated with the observed points $\xi_{N,j1}, \dots, \xi_{N,jn}$ (*n.b.* function $p_{i,N}$ depends on n , but we omit the index n to simplify notation). We define similarly:

$$p_{i,N}(\xi_{N,jn} | \xi_{N,j1:n-1}) = \frac{p_{i,N}(\xi_{N,j1:n})}{p_{i,N}(\xi_{N,j1:n-1})}.$$

For each $\alpha \in (0, 1)$, we denote by $\beta_{n,N}(\alpha)$ the miss probability of the NP test of level α when quantization is applied *i.e.*, the infimum of $\mathbb{P}_1[L_{n,N} < \gamma]$ w.r.t. all γ s.t. $\mathbb{P}_0[L_{n,N} > \gamma] \leq \alpha$. The error exponent associated with $\beta_{n,N}(\alpha)$ is provided by the following result, whose proof is similar to the one of Theorem 1.

Corollary 1: Consider a fixed $N \geq 2$. If Assumption 1 holds and if $(\log p_{i,N}(Z_{N,0}|Z_{N,-m:-1}))_{m \geq 0}$ is a convergent sequence in $L^1(\mathbb{P}_0)$ for each $i \in \{0, 1\}$ then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_{n,N}(\alpha) = -K_N,$$

where K_N is the constant defined by:

$$K_N = \lim_{m \rightarrow \infty} \mathbb{E}_0 \left[\log \frac{p_{0,N}}{p_{1,N}}(Z_{N,0}|Z_{N,-m:-1}) \right]. \quad (4)$$

The above result provides the error exponent K_N associated with the NP test on quantized observations. A natural question is: How does the choice of the quantizer affect the error exponent? Unfortunately, the expression of the error exponent does not immediately allow to evaluate the impact of the quantizer. In the sequel, we thus follow the approach of [6], [18] and focus on the case where the order N of the quantizer tends to infinity. We refer to such quantizers as *high-rate* quantizers. This approach leads to a convenient and informative asymptotic expression of K_N . In particular, it will be shown that, under some assumptions on the process $(Y_k)_{k \in \mathbb{Z}}$ and the quantizers sequence $(\mathcal{C}_N, \Xi_N, \xi_N)_{N \geq 1}$, the above error exponent K_N converges to K as N tends to infinity.

IV. PERFORMANCE OF HIGH-RATE VECTOR QUANTIZERS

A. Notation and Assumptions

For each N , we remark that the error exponent K_N does not depend on the particular choice of the quantization alphabet Ξ_N .² For the sake of simplicity, we assume with no loss of generality that³:

$$\xi_{N,j} = \frac{\int_{C_{N,j}} y \, dy}{\int_{C_{N,j}} dy} ,$$

i.e. each $\xi_{N,j}$ coincides with the centroid of cell $C_{N,j}$. We respectively define the volume and the diameter of cell j by $V_{N,j} = \int_{C_{N,j}} dy$ and $d_{N,j} = \sup_{u,v \in C_{N,j}} \|u - v\|$. We introduce the *specific point density* ζ_N and the *specific covariation profile* M_N as the piecewise constant functions on \mathcal{Y} respectively defined as follows, for any $y \in C_{N,j}$ ($j \in \{1, \dots, N\}$):

$$\begin{aligned} \zeta_N(y) &= \zeta_{N,j} = \frac{1}{NV_{N,j}} , \\ M_N(y) &= M_{N,j} = \frac{1}{V_{N,j}^{1+2/d}} \int_{C_{N,j}} (y - \xi_{N,j})(y - \xi_{N,j})^T dy . \end{aligned}$$

Now consider a family of quantizers $(\mathcal{C}_N, \Xi_N, \xi_N)_{N \geq 1}$. We make the following assumption.

Assumption 3: The following properties hold true.

- 1) As $N \rightarrow \infty$, ζ_N converges uniformly to a continuous function ζ such that $\inf_{y \in \mathcal{Y}} \zeta(y) > 0$.
- 2) As $N \rightarrow \infty$, M_N converges uniformly to a continuous (matrix-valued) function M such that $\sup_{y \in \mathcal{Y}} \|M(y)\| < \infty$.
- 3) There exists a constant C_d such that, for all N , $\sup_j d_{N,j} \leq \frac{C_d}{N^{1/d}}$.

We will refer to ζ as the *model point density* of the family $(\mathcal{C}_N, \Xi_N, \xi_N)_{N \geq 1}$. It represents the fraction of cells in the neighborhood of a given point y . Function M will be referred to as the *model covariation profile*. For each $y \in \mathcal{Y}$, $M(y)$ is a non-negative $d \times d$ matrix. In the literature, function $y \mapsto \text{Tr}(M(y))$ is usually referred to as the *inertial profile* [3], [6], [18]. Function M provides information about the shape of the cells.

²The value of the log-likelihood ratio (and *a fortiori* the value of the error exponent) remains unchanged by any one-to-one transformation of the quantized observations. Otherwise stated, the particular definition of the quantization alphabet has no impact on the corresponding Neyman-Pearson test provided that the latter quantization alphabet is composed by N distinct elements.

³The i th component of $\xi_{N,j}$ is defined as $\xi_{N,j}^{(i)} \triangleq \left(\int_{C_{N,j}} y^{(i)} dy \right) / \left(\int_{C_{N,j}} dy \right)$.

Intuitively, high-rate quantizers should be constructed in such a way that $\zeta(y)$ is large at those points y for which a fine quantization is essential to discriminate the two hypotheses. Theorem 2 below provides a more rigorous formulation of this intuition.

Remark 4: Assumption 3 is essentially the same as the one traditionally made in the *high-rate quantization* framework [3], [6], [18]. The main difference lies in Assumption 3-3): Usually, the volume of each cell vanishes at speed $1/N$ while the diameter tends to zero. Our assumption introduces a constraint on the speed of convergence of the sequence of diameters $\{d_{N,j}\}$, which ensures that cells shrink at the same speed ($1/N^{1/d}$) on each dimension. Assumption 3 is for instance valid for sequence of quantizers constructed as *companders* [3], [5]. Such quantizers write as the composition of an invertible function (the so-called *compressor*) and a uniform quantizer. Since [5], it is known that any scalar quantizer can be written as a compander. Under mild conditions on the compressor, it can be shown that any sequence of companders with a given fixed compressor satisfies Assumption 3 (in this case, the model point density ζ is fully determined by the first order derivative of the compressor). This point is discussed in Section IV-C.

B. Error Exponent in the High-Rate Regime

Before stating the main result, we need further assumptions. For each $m \geq 0$ and each $i \in \{0, 1\}$, define:

$$\begin{aligned} \eta_i(m) &= \sup_{m' \geq m} \mathbb{E}_0 |\log p_i(Y_0|Y_{-m:-1}) - \log p_i(Y_0|Y_{-m':-1})|, \\ \eta_{i,N}(m) &= \sup_{m' \geq m} \mathbb{E}_0 |\log p_{i,N}(Z_{N,0}|Z_{N,-m:-1}) - \log p_{i,N}(Z_{N,0}|Z_{N,-m':-1})| \end{aligned} \quad (5)$$

Note that we already assumed in Theorem 1 and Corollary 1 that sequences $\log p_i(Y_0|Y_{-m:-1})$ and $\log p_{i,N}(Z_{N,0}|Z_{N,-m:-1})$ converge in $L^1(\mathbb{P}_0)$ as $m \rightarrow \infty$, meaning that $\eta_i(m)$ and $\eta_{i,N}(m)$ tend to zero. Now coefficients $\eta_i(m)$ and $\eta_{i,N}(m)$ characterize the speed at which $\log p_i(Y_0|Y_{-m:-1})$ and $\log p_{i,N}(Z_{N,0}|Z_{N,-m:-1})$ converge to their limits. They are therefore related to the mixing property of processes $Y_{\mathbb{Z}}$ and $Z_{N,\mathbb{Z}}$ (this point is discussed below in Remark 8). In the sequel, we will need to ensure that these limits are reached fast enough (see Assumption 4-3) below).

Assumption 4: The following properties hold true.

- 1) For any $n \geq 1$, $y_{1:n} \mapsto p_i(y_{1:n})$ is of class C_3 on \mathcal{Y}^n .

- 2) $\sup_{\{n \geq 1, y_{1:n} \in Y^n, 1 \leq k, \ell, r \leq n, 1 \leq h, \bar{b}, \bar{j} \leq d\}} \left| \frac{\partial^3 \log p_i}{\partial y_k^{(h)} \partial y_\ell^{(\bar{b})} \partial y_r^{(\bar{j})}}(y_{1:n}) \right| < \infty$.
 3) There exist two constants $C_e, \epsilon > 0$ such that for each $i \in \{0, 1\}$, $N \geq 2$ and $m \geq 0$,

$$\max(\eta_i(m), \eta_{i,N}(m)) \leq \frac{C_e}{(1+m)^{6+\epsilon}} . \quad (6)$$

- 4) For each $i \in \{0, 1\}$, each integers m, m', k such that $-m' \leq -m \leq 0 \leq k$:

$$\mathbb{E}_0 \|\nabla_{y_0} \log p_i(Y_{0:k}|Y_{-m:-1}) - \nabla_{y_0} \log p_i(Y_{0:k}|Y_{-m':-1})\| \leq \varphi_m , \quad (7)$$

$$\mathbb{E}_0 \|\nabla_{y_0} \log p_i(Y_k|Y_{-m:k-1})\| \leq \psi_k , \quad (8)$$

where $\sum_k \varphi_k$ and $\sum_k \psi_k$ are convergent series.

Assumption 4 will be discussed in details at the end of the present subsection. Particular examples of processes satisfying the above assumption are provided in Section VI and in the numerical results of Section VII. We are now in position to state our main result. Recall that $p_0(y)$ is the pdf of Y_0 under \mathbb{P}_0 . Recall that K and K_N are the error exponents associated with the NP test in the absence and in the presence of quantization respectively, given by (3) and (4). Note that Assumption 4-3) implies that both sequences $\eta_i(m)$ and $\eta_{i,N}(m)$ tend to zero. This guarantees that under Assumption 1 the conclusions of Theorem 1 and Corollary 1 hold true *i.e.*, error exponents K and K_N do exist.

Theorem 2: Under Assumptions 1, 3, 4, the following statement holds true:

As N tends to infinity, $N^{2/d}(K - K_N)$ converges to a constant D_e given by

$$D_e = \frac{1}{2} \int \frac{p_0(y)F(y)}{\zeta(y)^{2/d}} dy , \quad (9)$$

where function F is given by

$$F(y) = \mathbb{E}_0 \left[\ell(Y_{\mathbb{Z}})^T M(Y_0) \ell(Y_{\mathbb{Z}}) \mid Y_0 = y \right] , \quad (10)$$

and random variable $\ell(Y_{\mathbb{Z}})$ is the limit in $L^2(\mathbb{P}_0)$ of sequence $\left(\nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:k}) \right)_{k \geq 0}$.

The proof of Theorem 2 is given in Section V.

Theorem 2 states that when the order of the quantizer tends to infinity, the error exponent K_N associated with the NP test converges at speed $N^{-2/d}$ to the error exponent K that one would have obtained in the absence of quantization. Loosely speaking, if $\beta_{n,N}(\alpha)$ represents the miss probability of the NP test of level α , the approximation

$$\beta_{n,N}(\alpha) \simeq e^{-n \left(K - \frac{D_e}{N^{2/d}} \right)} \quad (11)$$

holds when both the number n of sensors and the order N of quantization are large. Quantity D_e represents the (normalized) loss in error exponent between the quantized and the unquantized cases, in the high-rate quantization regime.

Note that Equation (9) resembles to Bennett's formula [5, Equation (1.6)] and its vector extension for r th-power distortion [6, Equation (7)].

Remark 5: As a first consequence of Theorem 2, under some assumptions on the process, *classical* quantizers as those produced in an MSE perspective will lead to error exponent K_N which converges to K as N tends to infinity, at speed $N^{-2/d}$ (see Equation (11) above).

Remark 6: The particular situation where measurements $(Y_k)_{k \geq 0}$ are i.i.d. under both hypotheses was studied by Gupta and Hero [18]. In this case, function $F(y)$ reduces to:

$$F(y) = \nabla \Lambda(y)^T M(y) \nabla \Lambda(y) ,$$

where $\Lambda(y) = \log \frac{p_0(y)}{p_1(y)}$ is the single sample LLR. Then, expression (9) of D_e is consistent with the results of Gupta and Hero (see in particular [18, Equation (20)]).

Note that we assume that each joint density $p_0(y_{1:n})$ and $p_1(y_{1:n})$ is of class C_3 on Y^n . Gupta and Hero's assumption is weaker, since they only assume that “*the densities are twice continuously differentiable on an open set of probability 1*” [18, page 1956]. In fact, we need conditions on the third derivatives of the logarithm of the densities in order to find relevant upper bounds of the Taylor-Lagrange remainders in the expansion of the *joint* densities $p_i(y_{-m:u})$ in the general case (see the detailed proof in Section V).

Remark 7: We now provide some insights on the meaning of Assumption 4 and on the class of stationary processes which satisfy the latter. Assumptions 4-1) and 4-2) are mild technical conditions on the smoothness of the pdf of the observations. They encompass a large family of stochastic processes and are generally simple to validate on a case-by-case basis. As explained above, Assumption 4-3) can be interpreted as a condition on the speed at which past observations are forgotten. Quantities $\eta_i(m)$ and $\eta_{i,N}(m)$ can be interpreted as conditional mixing coefficients associated with the unquantized and quantized processes $(Y_k)_k$ and $(Z_{N,k})_k$ respectively (see Remark 8 below). Past observations must be forgotten at least at a polynomial speed faster than m^6 . Assumption 4-4) can be interpreted similarly as a forgetting property, which no longer involves the logarithm of the density of the observations, but its derivative. For instance, Assumption 4 is simple to verify in case of short-dependent processes (such as moving average

processes for instance) provided that the density of the observation is smooth enough. A similar remark holds for a wide class of Markov chains. In this case, Assumption 4 essentially reduces to a smoothness assumption on the density of the transition kernel. More generally, we prove in Section VI that Assumption 4 holds for a wide class of hidden Markov models: We provide sufficient conditions on the transition kernel such that Assumption 4 holds. See also the numerical results in Section VII.

Remark 8: It is worth making some remarks on the link between Assumption 4 and standard mixing conditions used in the literature on mixing processes [26], [28], [29]. The mixing property which is the closest to our setting is related to the notion of ψ -mixing. For two σ -fields \mathcal{U} and \mathcal{V} , define the following coefficient [26], [28]:

$$\psi(\mathcal{U}, \mathcal{V}) = \sup_{\substack{U \in \mathcal{U}, V \in \mathcal{V} \\ \mathbb{P}(U) > 0, \mathbb{P}(V) > 0}} \left| 1 - \frac{\mathbb{P}(U \cap V)}{\mathbb{P}(U) \mathbb{P}(V)} \right|.$$

Recall that a stochastic process $Y_{\mathbb{Z}}$ is said to be ψ -mixing when the sequence of ψ -mixing coefficients $\psi(\sigma(Y_{n+1}), \sigma(Y_{-\infty:0}))$ converges to zero. The classical ψ -mixing condition traduces the fact that, loosely speaking, current samples at time n tend to become independent of past samples Y_0, Y_{-1}, \dots as n increases. In our case, we need to ensure that current samples become independent of past ones *conditionally to intermediate values* $Y_{1:n}$. Usual ψ -mixing coefficient do not fully allow to grasp this property. In [30], we introduced the following *conditional* ψ -mixing coefficient for σ -fields \mathcal{U} , \mathcal{V} and \mathcal{W} :

$$\bar{\psi}_i(\mathcal{U}, \mathcal{V} | \mathcal{W}) = \sup_{U \in \mathcal{U}, V \in \mathcal{V}} \text{ess sup} \left| 1 - \frac{\mathbb{P}_i(U \cap V | \mathcal{W})}{\mathbb{P}_i(U | \mathcal{W}) \mathbb{P}_i(V | \mathcal{W})} \right|$$

where the essential supremum is taken w.r.t. \mathbb{P}_0 and where we use the convention $0/0 = 1$. The above coefficient can be interpreted as a measure of dependence (under \mathbb{P}_i) between \mathcal{U} and \mathcal{V} conditionally to \mathcal{W} . In particular, it coincides with the traditional ψ -mixing coefficient $\psi(\mathcal{U}, \mathcal{V})$ when \mathcal{W} is taken to be the whole space $B(Y^{\mathbb{Z}})$ and $\mathbb{P} = \mathbb{P}_0$. For each $n \geq 1$, we further define $\bar{\psi}_i(n) = \bar{\psi}_i(\sigma(Y_{n+1}), \sigma(Y_{-\infty:0}) | \sigma(Y_{1:n}))$ and $\bar{\psi}_i(0) = \bar{\psi}_i(\sigma(Y_1), \sigma(Y_{-\infty:0}))$ when $n = 0$. There exists a close link between the above conditional mixing coefficients and the set of coefficients $\eta_i(m)$ defined in (5). In particular, Theorem 2 is valid when Assumption 4-2) is replaced by the assumption that sequences $\bar{\psi}_1(n)$ and $\bar{\psi}_{i,N}(n) = \bar{\psi}_i(\sigma(Z_{N,n+1}), \sigma(Z_{N,-\infty:0}) | \sigma(Z_{N,1:n}))$ converge to zero at speed $n^{6+\epsilon}$. We refer to [30] for details.

The asymptotic loss in error exponent D_e depends on the quantizer through its model point density ζ and its model covariation profile M . In the sequel, we study the values of these parameters which attenuate as much as possible the loss D_e .

C. Determination of Relevant High-Rate Quantizers: Scalar case ($d = 1$)

We first address the case where measurements $(Y_k)_{k \geq 0}$ are real-valued. Assume without much loss of generality that each cell is connected (cells are intervals) *i.e.*, the quantizer is regular [4]. In this case, a straightforward derivation leads to $M_N(y) = 1/12$ for each y and each N . Therefore, function F simplifies to:

$$\begin{aligned} F(y) &= \frac{1}{12} \mathbb{E}_0 \left[\ell(Y_{\mathbb{Z}})^2 \mid Y_0 = y \right] \\ &= \frac{1}{12} \lim_{k \rightarrow \infty} \mathbb{E}_0 \left[\left(\frac{\partial}{\partial y_0} \log \frac{p_0}{p_1}(Y_{-k:k}) \right)^2 \mid Y_0 = y \right]. \end{aligned}$$

Using Holder's inequality on (9), it is straightforward to prove the following result.

Corollary 2: Assume that $d = 1$ and that cells are intervals. The error exponent loss D_e is such that:

$$D_e \geq \frac{1}{2} \left(\int [p_0(y)F(y)]^{1/3} dy \right)^3, \quad (12)$$

where equality holds in (12) when the model point density coincides with:

$$\zeta(y) = \frac{[p_0(y)F(y)]^{1/3}}{\int [p_0(s)F(s)]^{1/3} ds}.$$

The above corollary provides the optimal high-rate quantization rule for the initial hypothesis testing problem. Note that expression (12) is quite similar to [31, Equation (15)] which gives “the minimum distortion resulting with optimum level spacing” in an MSE perspective.

Remark 9: In practice, N -point scalar quantizer achieving a given model point density ζ can be easily implemented by means of a compander. Recall that a compander is defined as the composition of an invertible continuous function ϕ (the so-called compressor) and a uniform quantizer [3], [5]. To that end, it is sufficient to define the compressor ϕ as the primitive of ζ on the observation space. For example, if \mathcal{Y} is the segment $[a, b] \subset \mathbb{R}$, define $\phi(x) = \int_a^x \zeta(t)dt$. Next the output of the compander is quantized using a uniform N -point quantizer on the interval $[0, 1]$. Under the assumption that ζ is a Lipschitz function, it is straightforward to show that the resulting sequence of quantizers satisfies Assumption 3 *i.e.*, that it achieves the model point density ζ .

D. Determination of Relevant High-Rate Quantizers: Vector case ($d \geq 2$)

In the vector case, the determination of optimal high-rate quantization rules implies the joint minimization of expression (9) w.r.t. both functions ζ and M . Unfortunately, as remarked in [3], [32], it is not known what functions M are allowable as covariation profiles. The determination of the set of admissible couples (ζ, M) is an open problem, which is beyond the scope of this paper.

However, when M is fixed, the point density ζ which minimizes D_e can be easily expressed as a function of M . Once again, this is a consequence of Holder's inequality:

$$D_e \geq \frac{1}{2} \left(\int [p_0(y)F(y)]^{\frac{d}{d+2}} dy \right)^{\frac{d+2}{d}},$$

where equality is achieved when the point density coincides with:

$$\zeta(y) = \frac{[p_0(y)F(y)]^{\frac{d}{d+2}}}{\int [p_0(s)F(s)]^{\frac{d}{d+2}} ds}. \quad (13)$$

In other words, one can easily provide the optimal high-rate quantization rule for a given limiting covariation profile. Following the approach of [18], we study two special cases of covariation profile:

1) *Congruent cells with minimum moment of inertia*: In this paragraph, we focus on congruent cells with minimum moment of inertia *i.e.*, we assume that

$$\forall y \in \mathcal{Y}, \quad M(y) = \nu I_d, \quad (14)$$

for some $\nu > 0$, where I_d represents the $d \times d$ identity matrix.

Recall that Gersho [33] made the now widely accepted conjecture that when N tends to infinity, most cells (*i.e.*, all the cells except those which are close to the boundary of the considered domain) of a d -dimensional MSE-optimal quantizer become congruent to some tessellating d -dimensional polytope H_d^* . In such a case, $M(y)$ is independent of y . Furthermore, Zamir and Feder [34, Lemma 1] proved that the cells of the MSE-optimal lattice quantizers become “closer” to balls *i.e.*, with minimum moment of inertia, as dimension d grows.

For quantizers with covariation profile given by (14), the optimal point density (13) becomes:

$$\zeta(y) = \frac{[p_0(y)\bar{F}(y)]^{\frac{d}{d+2}}}{\int [p_0(s)\bar{F}(s)]^{\frac{d}{d+2}} ds}, \quad (15)$$

where function \bar{F} is defined by

$$\begin{aligned}\bar{F}(y) &= \mathbb{E}_0 \left[\|\ell(Y_{\mathbb{Z}})\|^2 \mid Y_0 = y \right] \\ &= \lim_{k \rightarrow \infty} \mathbb{E}_0 \left[\left\| \nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:k}) \right\|^2 \mid Y_0 = y \right].\end{aligned}\quad (16)$$

Design Algorithm: In practice, one would like to design an N -point quantizer which point density approximately equals (15) for some finite N . This can be achieved by means of well-established algorithms, the most popular of them being the Linde-Buzo-Gray (LBG) algorithm [35]. This algorithm is an iterative method which computes a Voronoi tessellation, and yields an MSE-optimal N -point quantizer, from a training set of data of some pdf $p_0(y)$.

An (N -point) MSE-optimal quantizer for density $p_0(y)$ minimizes $\mathbb{E}_0 [\|Y_0 - \xi_N(Y_0)\|^2]$. As the number of quantization points N tends to infinity, such a quantizer has the following model point density [3], [6]:

$$\zeta_{MSE}(y) = \frac{p_0(y)^{\frac{d}{d+2}}}{\int p_0(s)^{\frac{d}{d+2}} ds} . \quad (17)$$

Comparing Equations (15) and (17), we deduce that the proposed quantizer, whose model point density ζ is given by Equation (15), can be obtained in practice by simply feeding the classical LBG algorithm with a training set of data of the following pdf:

$$q^*(y) = \frac{p_0(y)\bar{F}(y)}{\int p_0(s)\bar{F}(s) ds} .$$

Section VII provides numerical illustrations of this approach.

2) *Ellipsoidal cells:* In order to yield some insights on the general shape of the cells, and following [18], we focus in this paragraph on ellipsoidal cells. This kind of cells can not partition the considered convex subset \mathcal{Y} of \mathbb{R}^d but, for large dimension d , in analogy with the spherical cell approximation [3], [34], [36], we may assume that almost all cells of a given quantizer are close to ellipsoids.

Such an ellipsoidal cell, in the neighborhood of point y writes $C = \{x : (x-y)^T R(y) (x-y) \leq 1\}$, for some symmetric positive definite matrix $R(y)$. The corresponding covariation profile writes $M(y) = \nu |R(y)|^{1/d} R(y)^{-1}$ [18], [37], for some $\nu > 0$, and has an eigendecomposition

$$M(y) = U(y) \Phi(y) U(y)^T ,$$

where $\Phi(y) = \text{Diag}(\phi_{1:d}(y))$,⁴ and $U(y)$ is an orthogonal matrix. Note that the (positive) eigenvalues $(\phi_i(y))_{i \in \{1, \dots, d\}}$ of $M(y)$ capture the relative importance of the axes of the ellipsoid C , while the columns $(u_i(y))_{i \in \{1, \dots, d\}}$ of $U(y)$ *i.e.*, the eigenvectors of $M(y)$, indicate their respective directions.

In this paragraph, we assume that eigenvalues $(\phi_i)_{i \in \{1, \dots, d\}}$ are fixed, constant w.r.t. y and, without loss of generality, sorted in increasing order *i.e.*, $0 < \phi_1 \leq \phi_2 \leq \dots \leq \phi_d$. We want to find the best orthogonal matrix $U(y)$ *i.e.*, the one which minimizes function $F(y)$, given at Equation (10), in order to minimize the error exponent loss D_e (9). In other words, for a given “shape” of (non-degenerate) ellipsoid, we look for the best directions of its axes. Function $F(y)$ writes:

$$\begin{aligned} F(y) &= \mathbb{E}_0 \left[\ell(Y_{\mathbb{Z}})^T M(Y_0) \ell(Y_{\mathbb{Z}}) \mid Y_0 = y \right] \\ &= \text{Tr} \left(U(y) \Phi U(y)^T \bar{L}(y) \right) , \end{aligned} \quad (18)$$

where $\bar{L}(y) = \mathbb{E}_0 \left[\ell(Y_{\mathbb{Z}}) \ell(Y_{\mathbb{Z}})^T \mid Y_0 = y \right]$. Now write the eigendecomposition of the positive definite matrix $\bar{L}(y)$:

$$\bar{L}(y) = V(y) \Delta(y) V(y)^T ,$$

where $\Delta(y) = \text{Diag}(\lambda_{1:d}(y))$, $(\lambda_i(y))_{i \in \{1, \dots, d\}}$ are the (positive) eigenvalues of $\bar{L}(y)$ sorted in increasing order *i.e.*, $0 < \lambda_1(y) \leq \lambda_2(y) \leq \dots \leq \lambda_d(y)$, and $V(y)$ is an orthogonal matrix. Equation (18) thus writes:

$$\begin{aligned} F(y) &= \text{Tr} \left(U(y) \Phi U(y)^T V(y) \Delta(y) V(y)^T \right) \\ &\geq \sum_{i=1}^d \phi_i \lambda_{d-i+1}(y) , \end{aligned}$$

where the last inequality follows from a well-known trace inequality for positive semidefinite Hermitian matrices [38], [39, Section 9-H]. The above lower bound is furthermore achieved choosing matrix $U(y)$ such that $U(y)^T V(y)$ is the anti-diagonal matrix with ones on its anti-diagonal *i.e.*, defining the i th column of matrix $U(y)$ as the $(d-i+1)$ th column of matrix $V(y)$, or equivalently eigenvector $u_i(y)$ of matrix $M(y)$ as eigenvector $v_{d-i+1}(y)$ of matrix $\bar{L}(y)$.

⁴For any given d -dimensional vector $x_{1:d} \in \mathbb{R}^d$, $\text{Diag}(x_{1:d})$ represents the d -by- d diagonal matrix with diagonal coefficients (x_1, x_2, \dots, x_d) .

From the above derivations, we conclude that if a cell is a non-degenerate ellipsoid around y then its axes should be aligned along the ones of matrix $\bar{L}(y)$ in the reverse order. In particular, its minor axis should be aligned along the principal eigenvector of matrix $\bar{L}(y)$.

V. PROOF OF THEOREM 2

A. Preliminaries

Recall that $V_{N,j} = \int_{C_{N,j}} dy$ is the volume of cell $C_{N,j}$ ($j \in \{1, \dots, N\}$). For each $i \in \{0, 1\}$ and each set of quantization points $\xi_{N,j_{1:n}} = (\xi_{N,j_1}, \dots, \xi_{N,j_n}) \in \Xi_N^n$, define the following rescaled pdf of $Z_{N,1:n}$:

$$\begin{aligned} \bar{p}_{i,N}(\xi_{N,j_{1:n}}) &= \frac{1}{V_{N,j_1} \times \dots \times V_{N,j_n}} p_{i,N}(\xi_{N,j_{1:n}}) \\ &= \frac{1}{V_{N,j_1} \times \dots \times V_{N,j_n}} P_{i,n}(C_{N,j_1} \times \dots \times C_{N,j_n}) . \end{aligned} \quad (19)$$

The above definition is convenient because $\bar{p}_{i,N}(\xi_{N,j_{1:n}}) \simeq p_i(\xi_{N,j_{1:n}})$ for large values of N . This approximation will be of prime importance in the sequel. We define function $\bar{p}_{i,N}(\xi_{N,j_n} | \xi_{N,j_{1:n-1}})$ similarly.

For each $i \in \{0, 1\}$ and each integer $\ell \geq 0$, we introduce the following functions:

$$\begin{aligned} \forall y_{-\ell:0} \in Y^{\ell+1}, \quad \mathcal{L}_i(y_{-\ell:0}) &= \log p_i(y_0 | y_{-\ell:-1}) , \\ \forall z_{-\ell:0} \in \Xi_N^{\ell+1}, \quad \mathcal{L}_{i,N}(z_{-\ell:0}) &= \log \bar{p}_{i,N}(z_0 | z_{-\ell:-1}) . \end{aligned}$$

Due to Assumptions 1-3) and 4-3) (which ensures that $\eta_i(0) < \infty$), random sequence $(\mathcal{L}_i(Y_{-\ell:0}))_{\ell \geq 0}$ lies in $L^1(P_0)$. Moreover, Assumption 4-3) for large m ensures that sequence $(\mathcal{L}_i(Y_{-\ell:0}))_{\ell \geq 0}$ is a Cauchy sequence of $L^1(P_0)$. Denote by $\mathcal{L}_i(Y_{-\infty:0})$ its limit. From Assumption 4-3) once again, the following holds for any $\ell \geq 0$,

$$\mathbb{E}_0 |\mathcal{L}_i(Y_{-\ell:0}) - \mathcal{L}_i(Y_{-\infty:0})| \leq \frac{C_e}{(1 + \ell)^{6+\epsilon}} . \quad (20)$$

A similar result holds for sequence $(\mathcal{L}_{i,N}(Z_{N,-\ell:0}))_{\ell \geq 0}$ which converges in $L^1(\mathbb{P}_0)$ to some random variable $\mathcal{L}_{i,N}(Z_{N,-\infty:0})$ and verifies for any $\ell \geq 0$,

$$\mathbb{E}_0 |\mathcal{L}_{i,N}(Z_{N,-\ell:0}) - \mathcal{L}_{i,N}(Z_{N,-\infty:0})| \leq \frac{C_e}{(1 + \ell)^{6+\epsilon}} . \quad (21)$$

Our aim is to study the difference $K - K_N$ between error exponents associated with the ideal and quantized cases respectively. We may write the difference as

$$K - K_N = (K_0 - K_{0,N}) - (K_1 - K_{1,N}) , \quad (22)$$

where we defined for each $i \in \{0, 1\}$,

$$\begin{aligned} K_i &= \mathbb{E}_0 [\mathcal{L}_i(Y_{-\infty:0})] , \\ K_{i,N} &= \mathbb{E}_0 [\mathcal{L}_{i,N}(Z_{N,-\infty:0})] . \end{aligned}$$

In the sequel, we focus on the study of $K_1 - K_{1,N}$, the study of $K_0 - K_{0,N}$ being similar.

We now proceed with the proof. Choose any ϵ' such that $0 < \epsilon' < \frac{\epsilon}{3d(6+\epsilon)}$. Define the sequence of integers $m = m(N) = \lfloor N^{1/(3d)-\epsilon'} \rfloor$. We shall remember that with this definition,

$$\lim_{N \rightarrow \infty} \frac{m^3}{N^{1/d}} = 0 . \quad (23)$$

The following decomposition holds true: $K_{1,N} = K_1 + T_N + U_N + \delta_N$, where we defined:

$$\begin{aligned} T_N &= \mathbb{E}_0 [\mathcal{L}_{1,N}(Z_{N,-m:0}) - \mathcal{L}_1(Z_{N,-m:0})] , \\ U_N &= \mathbb{E}_0 [\mathcal{L}_1(Z_{N,-m:0}) - \mathcal{L}_1(Y_{-m:0})] , \\ \delta_N &= \mathbb{E}_0 [\mathcal{L}_{1,N}(Z_{N,-\infty:0}) - \mathcal{L}_{1,N}(Z_{N,-m:0})] + \mathbb{E}_0 [\mathcal{L}_1(Y_{-m:0}) - \mathcal{L}_1(Y_{-\infty:0})] . \end{aligned}$$

Using Equations (20) and (21), it is straightforward to show that

$$N^{2/d} |\delta_N| \leq 2 C_e \frac{N^{2/d}}{(1+m)^{6+\epsilon}} .$$

By definition of $m = m(N)$, we deduce that $N^{2/d} |\delta_N|$ converges to zero as $N \rightarrow \infty$. As a consequence, the asymptotic analysis of quantity $N^{2/d}(K_{1,N} - K_1)$ reduces to the study of T_N and U_N .

As \mathcal{Y} is a bounded set, Assumption 4-2) implies the following bounds on the derivatives of density p_1 which will be of permanent use in the sequel:

$$\sup_{\{y_{1:n} \in \mathcal{Y}^n, 1 \leq k \leq n\}} \|\nabla_{y_k} \log p_1(y_{1:n})\| \leq C_1 , \quad (24)$$

$$\sup_{\{y_{1:n} \in \mathcal{Y}^n, 1 \leq k \leq n\}} \|\nabla_{y_k}^2 \log p_1(y_{1:n})\| \leq C_2 , \quad (25)$$

for some constants C_1 and C_2 .

B. Study of T_N

We expand T_N as follows:

$$T_N = \mathbb{E}_0 \left[\log \frac{\bar{p}_{1,N}(Z_{N,-m:0})}{p_1(Z_{N,-m:0})} \right] - \mathbb{E}_0 \left[\log \frac{\bar{p}_{1,N}(Z_{N,-m:-1})}{p_1(Z_{N,-m:-1})} \right] . \quad (26)$$

We now study each term of the r.h.s. of the above equality. Consider $u \in \{-1, 0\}$. Writing the Taylor-Lagrange expansion of function $y_{-m:u} \mapsto p_1(y_{-m:u})$ at point $\xi_{N,j_{-m:u}}$, using Assumptions 3-3), 4 and the properties of the quantizers sequence, we prove the following lemma (the detailed proof is given in Appendix A).

Lemma 2: For each $j_{-m:u} \in \{1, \dots, N\}^{u+m+1}$, the following expansion holds true:

$$\frac{\bar{p}_{1,N}(\xi_{N,j_{-m:u}})}{p_1(\xi_{N,j_{-m:u}})} = 1 + \frac{1}{2N^{2/d}} \sum_{k=-m}^u \text{Tr} \left(\frac{\nabla_{y_k}^2 p_1(\xi_{N,j_{-m:u}})^T}{p_1(\xi_{N,j_{-m:u}})} \frac{M_{N,j_k}}{\zeta_{N,j_k}^{2/d}} \right) + \epsilon_{N,j_{-m:u}} ,$$

where $|\epsilon_{N,j_{-m:u}}| \leq c_T \left(\frac{m+1}{N^{1/d}} \right)^3$ for some constant c_T .

Plugging the above equation into (26), using $|\log(1+x) - x| \leq x^2$ in a neighborhood of zero, Assumptions 3, 4-2) and Equation (23), we obtain:

$$T_N = T_N(0) - T_N(-1) + o_N(N^{-2/d}) , \quad (27)$$

where, for each $u \in \{-1, 0\}$,

$$T_N(u) = \frac{1}{2N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\text{Tr} \left(\frac{\nabla_{y_k}^2 p_1(Z_{N,-m:u})^T}{p_1(Z_{N,-m:u})} \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \right) \right] . \quad (28)$$

C. Study of U_N

We expand U_N as follows:

$$U_N = \mathbb{E}_0 [\log p_1(Z_{N,-m:0}) - \log p_1(Y_{-m:0})] - \mathbb{E}_0 [\log p_1(Z_{N,-m:-1}) - \log p_1(Y_{-m:-1})] , \quad (29)$$

and study each term of the r.h.s. of the above equality. For each $u \in \{-1, 0\}$ and each $j_{-m:u} \in \{1, \dots, N\}^{u+m+1}$, we expand function $y_{-m:u} \mapsto \log p_1(y_{-m:u})$ at point $\xi_{N,j_{-m:u}}$:

$$\begin{aligned} \log p_1(y_{-m:u}) &= \log p_1(\xi_{N,j_{-m:u}}) + \sum_{k=-m}^u \nabla_{y_k} \log p_1(\xi_{N,j_{-m:u}})^T (y_k - \xi_{N,j_k}) \\ &\quad + \frac{1}{2} \sum_{k,\ell=-m}^u (y_k - \xi_{N,j_k})^T \nabla_{y_k, y_\ell}^2 \log p_1(\xi_{N,j_{-m:u}}) (y_\ell - \xi_{N,j_\ell}) + \epsilon'_N(y_{-m:u}) . \end{aligned} \quad (30)$$

Under Assumptions 3-3) and 4-2), for each $y_{-m:u} \in C_{N,j_{-m}} \times \dots \times C_{N,j_u}$, the remainder is such that

$$|\epsilon'_N(y_{-m:u})| \leq (m+1)^3 c'_3 \left(\frac{C_d}{N^{1/d}} \right)^3 ,$$

for some constant c'_3 . By Equation (23), the r.h.s. of the above inequality converges to zero as N tends to infinity faster than $N^{-2/d}$. Plugging Taylor expansion (30) into the expression (29) of U_N , we obtain:

$$U_N = U_N(0) - U_N(-1) + o_N(N^{-2/d}) , \quad (31)$$

where, for each $u \in \{-1, 0\}$,

$$U_N(u) = - \sum_{k=-m}^u \mathbb{E}_0 [\nabla_{y_k} \log p_1(Z_{N,-m:u})^\top (Y_k - Z_{N,k})] \\ - \frac{1}{2} \sum_{k,\ell=-m}^u \mathbb{E}_0 [(Y_k - Z_{N,k})^\top \nabla_{y_k, y_\ell}^2 \log p_1(Z_{N,-m:u}) (Y_\ell - Z_{N,\ell})] . \quad (32)$$

The next step is to study each dominant term of the r.h.s. of (32). The proof of the following lemma is provided in Appendix B.

Lemma 3: The following equality holds true for each $u \in \{-1, 0\}$:

$$U_N(u) = A_N(u) + B_N(u) + o_N(N^{-2/d}) ,$$

where A_N and B_N are defined as follows:

$$A_N(u) = - \frac{1}{N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_0(Y_{-m:u}) \right] , \quad (33) \\ B_N(u) = - \frac{1}{2N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\text{Tr} \left(\nabla_{y_k}^2 \log p_1(Z_{N,-m:u}) \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \right) \right] .$$

Now we expand the term $\nabla_{y_k}^2 \log p_1$ as follows:

$$\nabla_{y_k}^2 \log p_1(y_{-m:u}) = \frac{\nabla_{y_k}^2 p_1(y_{-m:u})}{p_1(y_{-m:u})} - \frac{\nabla_{y_k} p_1(y_{-m:u}) \nabla_{y_k} p_1(y_{-m:u})^\top}{(p_1(y_{-m:u}))^2} .$$

From the above decomposition and Equation (28), we can divide $B_N(u)$ into two terms:

$$B_N(u) = \frac{1}{2N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\text{Tr} \left(\nabla_{y_k} \log p_1(Z_{N,-m:u}) \nabla_{y_k} \log p_1(Z_{N,-m:u})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \right) \right] - T_N(u) .$$

Expanding function $\nabla_{y_k} \log p_1$ in the above equation and in (33), we can write dominant terms in a simple form *i.e.*, replace each Z_N by Y . Under Assumption 3, from Equations (25) and (23), we can easily prove that the corresponding remainders are $o_N(N^{-2/d})$.

Putting all pieces together, we obtain

$$\begin{aligned}
U_N(u) = & -\frac{1}{N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Y_{-m:u})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_0(Y_{-m:u}) \right] \\
& + \frac{1}{2N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Y_{-m:u})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_1(Y_{-m:u}) \right] \\
& - T_N(u) + o_N(N^{-2/d}) . \quad (34)
\end{aligned}$$

D. End of the Proof

From the results of sections V-B and V-C, we can easily prove the following lemma.

Lemma 4: The following holds true:

$$N^{2/d}(K - K_N) = \mathbb{E}_0[\mathcal{H}_{N,0}(Y_{-m:0})] + \sum_{k=-m}^{-1} \mathbb{E}_0[\mathcal{H}_{N,k}(Y_{-m:0}) - \mathcal{H}_{N,k}(Y_{-m:-1})] + o_N(1) , \quad (35)$$

where for each $u \in \{-1, 0\}$, each $m \geq 1$ and each $k \in \{-m, \dots, u\}$:

$$\mathcal{H}_{N,k}(Y_{-m:u}) = \frac{1}{2} \nabla_{y_k} \log \frac{p_0}{p_1}(Y_{-m:u})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log \frac{p_0}{p_1}(Y_{-m:u}) . \quad (36)$$

Proof: Recalling the decomposition: $K_{1,N} = K_1 + T_N + U_N + o_N(N^{-2/d})$ and gathering Equations (27), (31), (34), it is straightforward to prove the following equality:

$$\begin{aligned}
N^{2/d}(K_{1,N} - K_1) = & - \sum_{k=-m}^0 \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Y_{-m:0})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_0(Y_{-m:0}) \right] \\
& + \frac{1}{2} \sum_{k=-m}^0 \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Y_{-m:0})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_1(Y_{-m:0}) \right] \\
& + \sum_{k=-m}^{-1} \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Y_{-m:-1})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_0(Y_{-m:-1}) \right] \\
& - \frac{1}{2} \sum_{k=-m}^{-1} \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Y_{-m:-1})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \nabla_{y_k} \log p_1(Y_{-m:-1}) \right] \\
& + o_N(1) .
\end{aligned}$$

Similar expression holds for $N^{2/d}(K_{0,N} - K_0)$ —replace all p_1 by p_0 in the above equation. Lemma 4 follows from decomposition (22). ■

We now study the series (35). From Assumptions 3, 4-2) and 4-4), the following forgetting properties hold true for any positive integers ℓ' , ℓ and any integers k , u s.t. $-\ell' \leq -\ell \leq k \leq u$:

$$\mathbb{E}_0 |\mathcal{H}_{N,k}(Y_{-\ell:u}) - \mathcal{H}_{N,k}(Y_{-\ell':u})| \leq c_h \varphi_{\ell-|k|} , \quad (37)$$

$$\mathbb{E}_0 |\mathcal{H}_{N,k}(Y_{-\ell:0}) - \mathcal{H}_{N,k}(Y_{-\ell:-1})| \leq c_h \psi_{|k|} , \quad (38)$$

for some constant c_h .

It is clear from (37) that sequence $(\mathcal{H}_{N,k}(Y_{-\ell:u}))_{\ell \geq -u}$ is a Cauchy sequence in $L^1(\mathbb{P}_0)$. We simply denote its limit by $\mathcal{H}_{N,k}(Y_{-\infty:u})$. Inequalities (37) and (38) provide the main tools for the asymptotic analysis of series (35). The proof of the following lemma is given in Appendix C.

Lemma 5: The following holds true:

$$N^{2/d}(K - K_N) = \mathbb{E}_0 [\mathcal{H}_{N,0}(Y_{-\infty:0})] + \sum_{k=-\infty}^{-1} \mathbb{E}_0 [\mathcal{H}_{N,k}(Y_{-\infty:0}) - \mathcal{H}_{N,k}(Y_{-\infty:-1})] + o_N(1) .$$

As process $(Y_k)_{k \in \mathbb{Z}}$ is stationary, the expectation \mathbb{E}_0 enclosed in the sum of the above equation is invariant w.r.t. a time-shift. Using this remark, we obtain after algebra

$$N^{2/d}(K - K_N) = \lim_{k \rightarrow \infty} \mathbb{E}_0 [\mathcal{H}_{N,0}(Y_{-\infty:k})] + o_N(1) . \quad (39)$$

For a fixed $k \geq 0$, Equation (7) ensures that sequence $\left(\nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-m:k}) \right)_{m \geq 0}$ is a Cauchy sequence in $L^1(\mathbb{P}_0)$. Denote its limit by $\ell_k(Y_{-\infty:k})$. The upper bound of Equation (8) is uniform in m . Consequently, it also holds for sequence $(\ell_k(Y_{-\infty:k}))_{k \geq 0}$:

$$\mathbb{E}_0 \|\ell_k(Y_{-\infty:k}) - \ell_{k-1}(Y_{-\infty:k-1})\| \leq \psi_k .$$

Under Assumption 4-4), $\sum_k \psi_k$ is a convergent series. Sequence $(\ell_k(Y_{-\infty:k}))_{k \geq 0}$ is thus a Cauchy sequence in $L^1(\mathbb{P}_0)$. Denote its limit by $\ell(Y_{\mathbb{Z}})$. Moreover, the upper bound of Equation (7) (resp. Equation (8)) is uniform in m' (resp. m). It is then straightforward to prove that $\ell(Y_{\mathbb{Z}})$ coincides with the $L^1(\mathbb{P}_0)$ -limit of sequence $\left(\nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:k}) \right)_{k \geq 0}$.

From Equation (24) and its counterpart for density p_0 , quantity $\nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:k})$ is uniformly bounded. Consequently, the above limit also holds in the $L^2(\mathbb{P}_0)$ -sense:

$$\nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:k}) \xrightarrow[k \rightarrow \infty]{L^2(\mathbb{P}_0)} \ell(Y_{\mathbb{Z}}) . \quad (40)$$

Plugging Equations (36) and (40) in Equation (39) and letting N tend to ∞ complete the proof of Theorem 2.

VI. ILLUSTRATION: CASE OF A HIDDEN MARKOV PROCESS

In this section, we translate our assumptions in the case of (discrete-time) hidden Markov models. For such models, they reduce to simpler conditions on the transition kernel of the underlying Markov chain, and on the observation kernel. This context, where the measurements are noisy samples of a certain Markov source, has raised a deep interest in the recent literature on sensor networks (see [23], [24] and reference therein).

Consider a stationary Markov process $(X_k)_{k \geq 0}$ taking its values in an arbitrary state space X , and playing the role of a source signal to be detected. For each $i \in \{0, 1\}$ and each integer t , we assume that the (iterated) transition kernel $\mathbb{P}_i[X_{k+t} \in \cdot | X_k = x]$ admits a density $x' \mapsto q_i^t(x, x')$ w.r.t. some probability measure λ on $(X, B(X))$. Assume that there exist an integer m , and two real numbers σ^-, σ^+ s.t., for each $i \in \{0, 1\}$ and each $(x, x') \in X^2$, $0 < \sigma^- \leq q_i^m(x, x') \leq \sigma^+$. In particular, this assumption implies that the Markov chain $(X_k)_{k \in \mathbb{Z}}$ has bounded support.

If the state space X is finite, the above conditions hold if the Markov chain $(X_k)_{k \in \mathbb{Z}}$ is *irreducible aperiodic*, choosing λ as the (normalized) counting measure on X . In this case, the chain indeed admits a stationary distribution, and $q_i^m(x, x') > 0$ for each x, x' and some integer m [40, Section 8].

The states X_k of the above Markov source are supposed to be hidden. However, a “noisy” version Y_k ($\in Y \subset \mathbb{R}^d$) of X_k is available at the k th sensor. We assume that the distribution $\mathbb{P}[Y_k \in \cdot | X_k = x]$ does not depend on the hypothesis H_0 or H_1 , and admits a density $y \mapsto g(x, y)$ w.r.t. the d -dimensional Lebesgue measure μ restricted to Y , such that $0 < \inf_{x,y} g(x, y) \leq \sup_{x,y} g(x, y) < \infty$. We furthermore assume that this density verifies some smoothness conditions: For each $x \in X$, $y \mapsto g(x, y)$ is of class C_3 on Y , and $\sup_{\{x \in X, y \in Y, 1 \leq h, \bar{h}, \bar{j} \leq d\}} \left| \frac{\partial^3 g}{\partial y^{(h)} \partial y^{(\bar{h})} \partial y^{(\bar{j})}}(x, y) \right| < \infty$. The situation is depicted in Figure 1.

A similar assumption was recently introduced by [41], [42] in order to study the asymptotic behaviour of the log-likelihood $\log p_i(Y_{1:n})$ as n tends to infinity. In particular, it was shown that:

$$|\log p_i(Y_0 | Y_{-m:-1}) - \log p_i(Y_0 | Y_{-m':-1})| \leq \frac{2}{1 - \sigma^- / \sigma^+} \left(\frac{\sigma^-}{\sigma^+} \right)^{m-1}$$

for each $m' \geq m \geq 0$. This clearly proves that sequence $\log p_i(Y_0 | Y_{-m:-1})$ converges in $L^1(\mathbb{P}_0)$ as $m \rightarrow \infty$ and yields Assumption 2. Moreover, the convergence holds at exponential speed, meaning that quantities $\eta_i(m)$, defined by Equation (5), vanish faster than $1/m^6$. The same claim

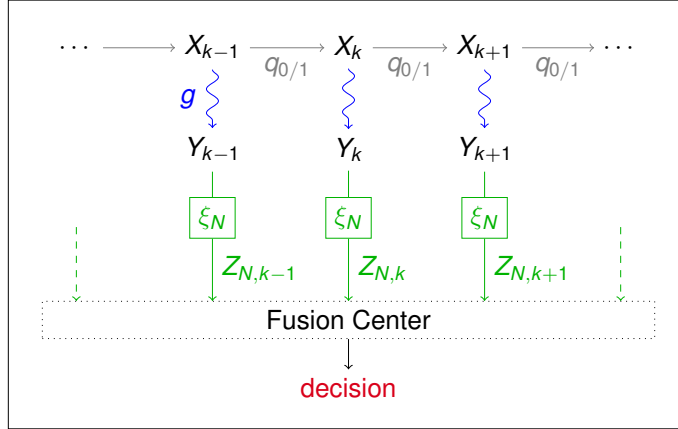


Figure 1: Detection of a discrete-time Markov process based on noisy observations.

holds as well for quantities $\eta_{i,N}(m)$, without need for any special condition on the quantizer (quantization preserves the hidden Markov nature of the original process $(Y_k)_{k \in \mathbb{Z}}$). This yields Assumption 4-3).

Assumptions 4-1) and 4-2) are direct consequences of the above smoothness conditions on density g . Assumption 4-4) can be derived following the arguments of [41], [42].

The following proposition then follows from the results of [41], [42]. The proof is therefore omitted.

Proposition 1: All conditions given by Assumptions 1 and 4 hold true for the particular process $(Y_k)_{k \in \mathbb{Z}}$ described in this section.

As a consequence, if the family of quantizers moreover verifies Assumption 3, then the conclusions of Theorems 1 and 2 hold true.

Section VII-A below provides a practical example of such a detection problem.

VII. NUMERICAL RESULTS

In this section, we provide numerical illustrations of the proposed quantization rule in terms of geometric properties and performance. Different contexts are considered and we compare several quantizers:

- The *proposed quantizer*, obtained using the approach described in Section IV-D1 and whose model point density is given by (15).

- The *MSE-optimal quantizer*, which minimizes $\mathbb{E}_0 \|Y_0 - Z_{N,0}\|^2$ and whose model point density is given by (17).
- *Gupta-Hero quantizer*, introduced in [18]: In this case the model point density is drawn as if observations were i.i.d. *i.e.*, only taking the marginal distributions $p_0(y)$ and $p_1(y)$ into account.
- The *uniform quantizer* with constant model point density.

A. Scenario #1: Detection of Quaternary Modulations: QPSK vs. OQPSK

In this section, we provide an example of hidden Markov models which verify the assumptions given at Section VI, and detail how to use in this case the approach described in Section IV-D1 for the design of practical quantizers.

1) *Observation Model*: We consider the following model for vector observations with dimension $d = 2$:

$$Y_k = T(X_k) + W_k, \quad (41)$$

where $(X_k)_{k \in \mathbb{Z}}$ is a 2-bit message, which takes values in $\mathcal{X} = \{0, 1, 2, 3\}$, $T(x)$ is the 2-D representation of state x in the I-Q plane⁵ according to Figure 2, and $W_k \stackrel{i.i.d.}{\sim} \mathcal{CN}(0, \sigma^2)$ represents a zero mean circular Gaussian thermal noise with variance σ^2 . Process $(X_k)_{k \in \mathbb{Z}}$ is i.i.d., uniformly distributed under H0, and forms a Markov chain under H1. More precisely,

$$\text{H0 : } X_k \stackrel{i.i.d.}{\sim} \mathcal{U}_{\{0,1,2,3\}}$$

$$\text{H1 : } X_0 \sim \mathcal{U}_{\{0,1,2,3\}}, \mathbb{P}_1[X_{k+1} = x' | X_k = x] = q(x, x'),$$

where q is the transition matrix of the Markov chain and is given by:

$$q = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{bmatrix}.$$

This situation arises when testing from noisy observations between two possible quaternary modulations, namely quadrature phase-shift keying (QPSK) and offset quadrature phase-shift keying (OQPSK), in the In-phase/Quadrature plane [43, Chapter 3]. The corresponding constellations are depicted in Figure 2.

⁵ $T(0) = [-1; -1]$, $T(1) = [-1; 1]$, $T(2) = [1; 1]$, $T(3) = [1; -1]$.

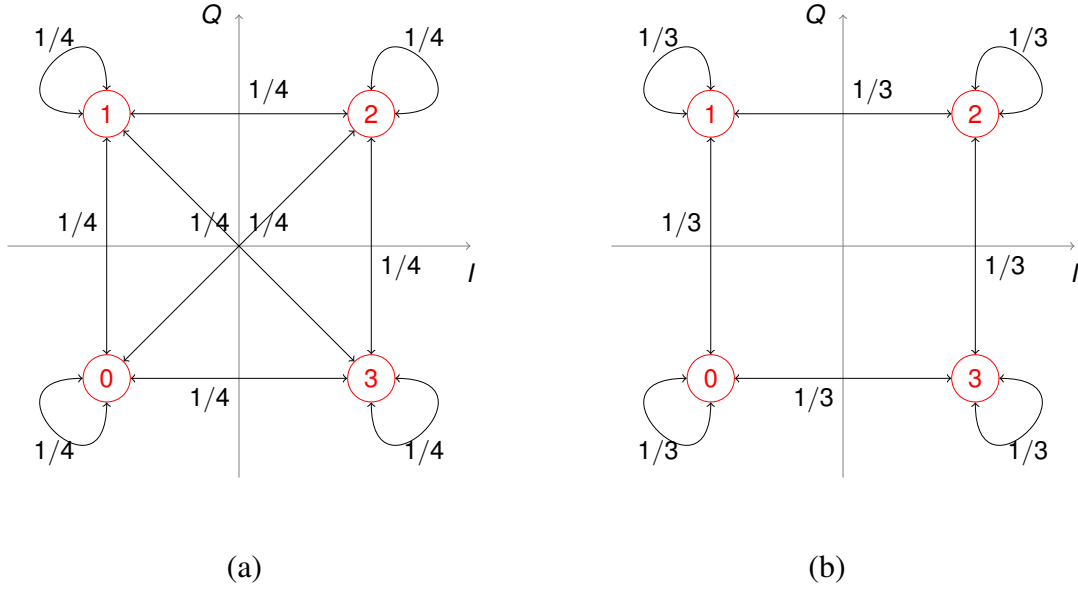


Figure 2: QPSK vs. OQPSK – Constellation diagrams and transitions probabilities for (a) QPSK, (b) OQPSK.

In the observation model (41), densities have infinite support. We thus consider truncated observations on $\mathcal{Y} = [-M; M]^2$ for some positive real number M [44, Section 10.1]. The new (truncated) model is a hidden Markov model with observation density $g(x, y)$ given by:

$$g(x, y) = \frac{\mathbf{1}_{[-M; M]^2}(y)}{C_M(\sigma)} \exp \left(\frac{-1}{2\sigma^2} (y - T(x))^T (y - T(x)) \right), \quad (42)$$

where $\mathbf{1}_A$ stands for the indicator function of set A , and $C_M(\sigma)$ is a constant such that $\int_{\mathcal{Y}} g(x, y) dy = 1$, for each $x \in \{0, 1, 2, 3\}$ i.e., $C_M(\sigma) = \left(\int_{-M}^M \exp \left(\frac{-(t-1)^2}{2\sigma^2} \right) dt \right)^2$.

The above hidden Markov model verifies the assumptions given at Section VI. From Proposition 1, if the family of quantizers verifies Assumption 3, then the conclusions of Theorems 1 and 2 hold true.

Note that the marginal pdf of the measurements $(Y_k)_{k \geq 0}$ (represented in Figure 3) writes

$$p_0(y) = p_1(y) = \frac{1}{4} \sum_{x=0}^3 g(x, y). \quad (43)$$

Since it does not depend on the hypothesis, Gupta-Hero quantizer [18], which minimizes the error exponent loss in case of i.i.d. observations, is not defined.

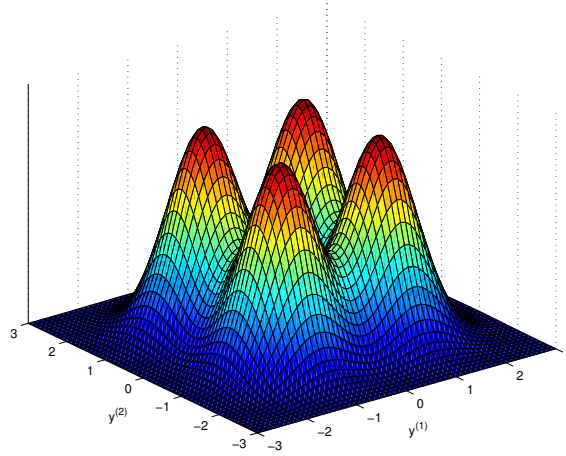


Figure 3: QPSK vs. OQPSK – Marginal pdf of the observations $p_0(y) = p_1(y)$ ($M = 3$, $\sigma = 0.6$).

2) *Examples of Quantizers:* Figure 4(a) represents the MSE-optimal 128-cell quantizer obtained by the LBG algorithm, and setting $M = 3$, $\sigma = 0.6$. Figure 4(b) represents the corresponding proposed quantizer. Our quantizer is significantly different from the MSE-optimal one. Some low probability points turn out to be significant for the considered detection problem. Details on how we obtained these quantizers are given below.

a) *MSE-optimal quantizer:* The MSE-optimal quantizer of Figure 4(a) was obtained by feeding the LBG algorithm with 20 000 samples following distribution \mathbb{P}_0 i.e., i.i.d. with pdf $p_0(y)$ (see Figure 3).

b) *Proposed quantizer:* As noted in Section IV-D1, the proposed quantizer, whose model point density ζ is given by Equation (15), can be obtained by simply feeding the LBG algorithm with observations corresponding with the following pdf:

$$q^*(y) = \frac{p_0(y)\bar{F}(y)}{\int p_0(s)\bar{F}(s) ds}.$$

We simulated 20 000 samples of this pdf using rejection sampling [45, Section 2.2]. In practice, we approximated function \bar{F} given by Equation (16) by:

$$\bar{F}_k(y) = \frac{1}{n_{MC}} \sum_{j=1}^{n_{MC}} \left\| \nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:-1}(j), y, Y_{1:k}(j)) \right\|^2, \quad (44)$$

for $k = 3$ and $n_{MC} = 1\,000$ replications $(Y_m(j))_{m \in \{-k, \dots, -1, 1, \dots, k\}, j \in \{1, \dots, n_{MC}\}}$ i.e., 6 000 i.i.d. samples with pdf p_0 . These values were chosen based on empirical observations.

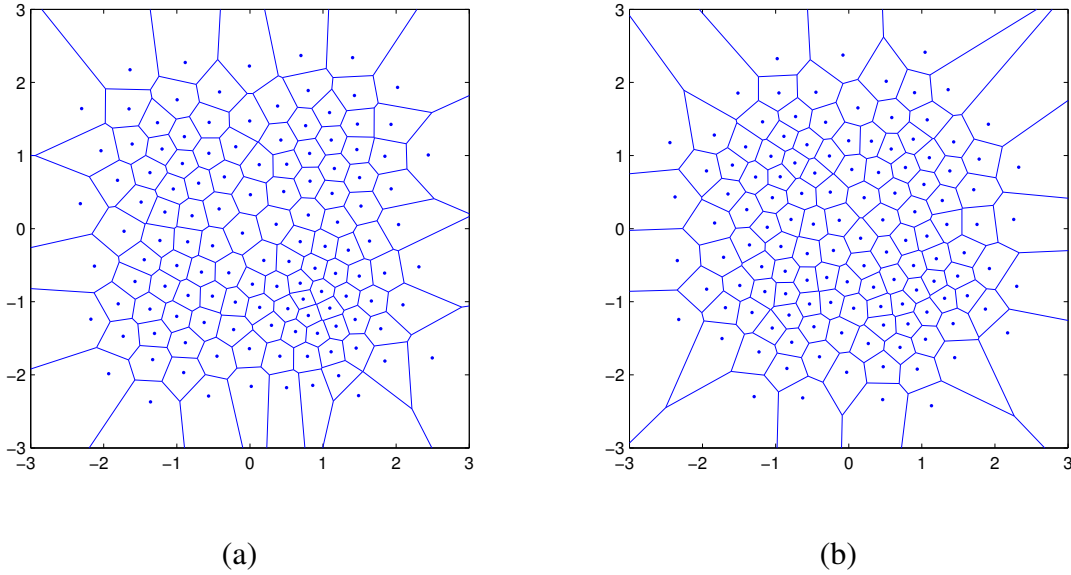


Figure 4: QPSK vs. OQPSK – (a) MSE-optimal 128-cell quantizer, (b) Proposed 128-cell quantizer ($M = 3$, $\sigma = 0.6$, 20 000 samples).

The gradient in the above equation may be written as follows, after some derivations, and using Equations (42), (43):

$$\begin{aligned}
 \nabla_{y_0} \log \frac{p_0}{p_1}(y_{-k:k}) &= \nabla_{y_0} \log p_0(y_0) - \nabla_{y_0} \log p_1(y_{-k:k}) \\
 &= \frac{1}{\sigma^2} \left\{ \frac{\mathbb{E}_0 [T(X_0) g(X_0, y_0)]}{\mathbb{E}_0 [g(X_0, y_0)]} - \frac{\mathbb{E}_1 [T(X_0) \prod_{j=-k}^k g(X_j, y_j)]}{\mathbb{E}_1 [\prod_{j=-k}^k g(X_j, y_j)]} \right\}.
 \end{aligned}$$

As they are finite sums on \mathbf{X} or \mathbf{X}^{2k+1} , the above four expectations are exactly computed at the time of the evaluation of \bar{F}_k (44).

B. Scenario #2: Detection of an AR Structure in Gaussian 2-D Signals

We consider the following model for vector observations with dimension $d = 2$:

$$Y_k = X_k + W_k,$$

where $W_k \stackrel{i.i.d.}{\sim} \mathcal{CN}(0, \sigma^2)$ represents a zero mean circular Gaussian thermal noise with variance σ^2 , and where $(X_k)_{k \in \mathbb{Z}}$ is a Gaussian process which is white under H_0 and correlated (AR-1)

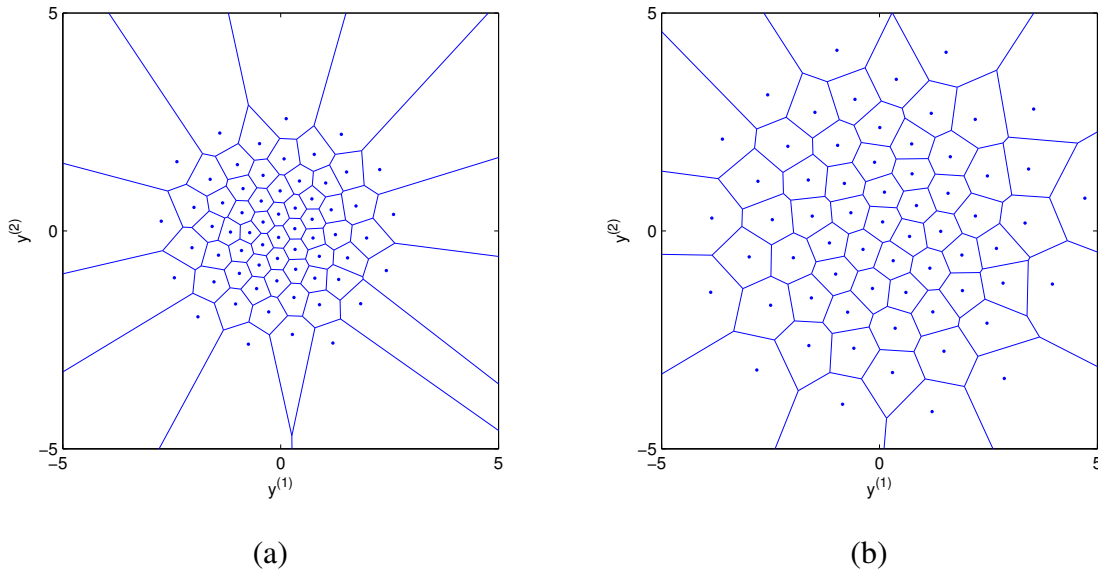


Figure 5: Detection of an AR structure – (a) MSE-optimal 64-cell quantizer, (b) Proposed 64-cell quantizer ($a = 0.8$, $\sigma = 1$, 20 000 samples).

under H1. More precisely,

$$\begin{aligned} H0 : X_k &\stackrel{i.i.d.}{\sim} \mathcal{CN}(0, 1) \\ H1 : X_k &= aX_{k-1} + \sqrt{1 - a^2} U_k , \end{aligned}$$

where $a \in (0, 1)$ is the correlation coefficient and $U_k \stackrel{i.i.d.}{\sim} \mathcal{CN}(0, 1)$ is the innovation process. In particular, $(Y_k)_{k \in \mathbb{Z}}$ is a white Gaussian process under H0 and is a hidden Markov process under H1, with the particular property that marginal distribution of single observations are identical under both hypotheses.

We mention that in the above model, densities have infinite support so that the assumptions made in this paper are not satisfied (the observation set \mathcal{Y} coincides with \mathbb{R}^2 and is thus unbounded). In particular, Theorem 2 does not apply. Nevertheless, in order to yield some insights on the design of practical quantizers for detection, we can still use the approach described in Section IV-D1 and compute the proposed model point density given by Equation (15).

Figure 5(a) represents the MSE-optimal 64-cell quantizer obtained by the LBG algorithm (with a 20 000-sample training set of data), and setting $\sigma = 1$. Figure 5(b) represents the corresponding

proposed quantizer⁶, obtained when setting $a = 0.8$. Once again, our quantizer is significantly different from the MSE-optimal one. As a matter of fact, low probability points seem to be significant for the considered detection problem.

Table I compares the latter two quantization rules and the uniform one (on the rectangle $[-8; 8]^2$) in terms of quantity D_e (9). As expected, the proposed quantization rule leads to the lowest one. We can guess it will also lead to higher detection performance.

Table I: Detection of an AR structure – Quantity D_e for parameters values $a = 0.8$ and $\sigma = 1$.

Quantization rule	Uniform on $[-8; 8]^2$	MSE-optimal	Proposed one
Quantity D_e	8.211	2.255	2.112

C. Scenario #3: Detection of a Scalar MA Process in Noise

Denote by Y_k the samples collected by a receiver which makes a binary test associated with the following hypotheses:

$$\begin{aligned} H0 : Y_k &= W_k , \\ H1 : Y_k &= \sum_{\ell=0}^L h_{\ell} U_{k-\ell} + W_k . \end{aligned}$$

where $W_k \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ represents a thermal noise which is supposed to be real-valued for the sake of illustration. Here, U_k represents a certain random source which is passed through a propagation channel with deterministic real coefficients h_0, \dots, h_L , where L is an integer which represents the channel's memory. In the sequel, we set $L = 3$. Assume for instance that U_k is Gaussian distributed $U_k \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. We investigate the case where the sensing unit performs a scalar quantization of the received signal before transmission to the decision device.

As in Section VII-B, in the above model, densities have infinite support so that the assumptions made in this paper are not satisfied. Once again, in order to yield some insights on the design of practical quantizers for detection, we can still use the approach described in Section IV-D1 and compute the proposed model point density given by Equation (15)⁷.

⁶In this case, we approximated function \bar{F} (16) for finite k and exactly computed the involved expectation.

⁷In this case, we approximated function \bar{F} (16) for finite k and exactly computed the involved expectation.

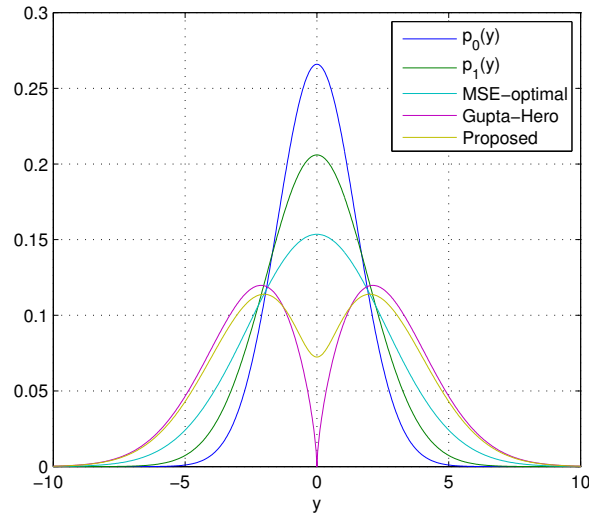


Figure 6: Detection of an MA process – Probability and model point densities ($h = [1.06677, -0.59281, 0.09565]$, $\sigma = 1.5$).

For the same reason, the result of Gupta and Hero [18, Equation (20)] does not apply, but we can compute the corresponding quantizer, which model point density is given by [18, Equation (25)], as they did for their Gaussian examples in [18, Section V].

The performance depend on the noise variance σ^2 and on the particular value of the channel. Thus, we assumed that channel coefficients h_0, \dots, h_L are i.i.d. Gaussian distributed with zero mean and unit variance, and made several simulations.

Figure 6 represents the probability and model point densities for one channel realization *i.e.*, $h = [1.06677, -0.59281, 0.09565]$, and setting $\sigma = 1.5$.

Considering a system with $n = 80$ sensors, constructing 4-cell quantizers for different methods, and computing the corresponding quantized probability distributions under each hypothesis, we can compare the considered quantization rules in terms of detection performance through their respective receiver operating characteristics (ROC curves). Figure 7 represents such curves for the above channel realization. The uniform quantizer is used on the support $[-10\sigma, 10\sigma]$. The whole curve is plotted using 50 000 samples of LLR under each hypothesis.

The proposed quantization rule improves the detection performance compared to the MSE-optimal quantizer. In this example, the ROC curve is close to that obtained using Gupta-Hero

quantizer. Recall however that in other contexts (*e.g.* in Scenarios #1 and #2), Gupta-Hero quantizer may not even be defined. We must also qualify this observation: Our theoretical results are valid in the asymptotic regime where N and n tend to infinity, that is, in the regime where the power of the test tends exponentially to one. In practice, the empirical validation of our result would thus require to simulate rare events. This topic is out of the scope of this paper.

Note that if we interchange H_0 and H_1 , the proposed quantization rule will be different. This is due to the fact that the asymptotic regime we are interested in when dealing with error exponents *i.e.*, n tends to infinity for a fixed type-I error α , restricts attention to one point along the Neyman-Pearson ROC curve.

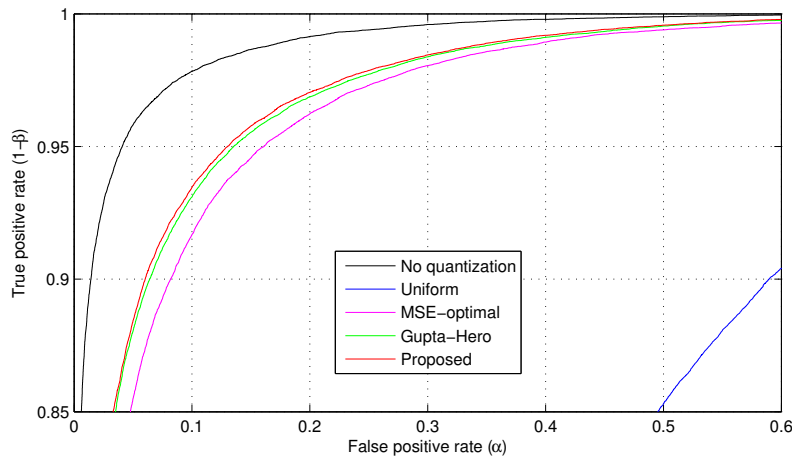


Figure 7: Detection of an MA process – ROC curves ($h = [1.06677, -0.59281, 0.09565]$, $\sigma = 1.5$, $n = 80$, $N = 4$, 100 000 samples).

VIII. CONCLUSION

We investigated the performance of the Neyman-Pearson detector used on quantized versions of a correlated vector-valued stationary process. It was shown that for a constant false alarm level, the miss probability of the test converges exponentially to zero. We determined the error exponent and we provided a compact and informative expression of the latter in the context of high-rate quantization. It is proved in particular that when the number N of quantization levels tends to infinity, the error exponent converges at speed $N^{-2/d}$ to the ideal error exponent that

one would obtain in the absence of quantization. In case of scalar quantization, we analytically characterized the high-rate quantizers minimizing the error exponent loss. In case of vector quantization, we proposed a method based on the LBG algorithm in order to construct practical quantizers with attractive performance.

We believe that there are many directions for extending these results and mention a few here. In this paper, observations have absolutely continuous probability distributions w.r.t. the Lebesgue measure. Following Graf and Luschgy [46, Section 6] who considered measures with both continuous and singular parts, we could think of an extension of our work to such cases.

We moreover focused on constant false-alarm rate (CFAR) tests. Following the arguments developed in [18] and using the results of [25, Section III], it could be interesting to study the whole asymptotic ROC curve and use a global performance criterion like the area under the curve (AUC). However, this would require a nontrivial extension of Sanov's theorem [47] to non-i.i.d. times series.

We furthermore think that the framework developed in this paper could be applied in the context of parameter estimation. The effect of quantization on performance, measured for instance by the Fisher information, could be studied and corresponding optimal vector quantizers could be described.

APPENDIX A

PROOF OF LEMMA 2

We write the Taylor-Lagrange expansion of function $y_{-m:u} \mapsto p_1(y_{-m:u})$ at point $\xi_{N,j-m:u}$:

$$\begin{aligned} p_1(y_{-m:u}) &= p_1(\xi_{N,j-m:u}) + \sum_{k=-m}^u \nabla_{y_k} p_1(\xi_{N,j-m:u})^T (y_k - \xi_{N,j_k}) \\ &\quad + \frac{1}{2} \sum_{k,\ell=-m}^u (y_k - \xi_{N,j_k})^T \nabla_{y_k, y_\ell}^2 p_1(\xi_{N,j-m:u}) (y_\ell - \xi_{N,j_\ell}) + \epsilon_N(y_{-m:u}) , \end{aligned} \quad (45)$$

where

$$\begin{aligned} \epsilon_N(y_{-m:u}) &= \frac{1}{6} \sum_{k,\ell,r=-m}^u \sum_{h,\bar{B},\bar{J}=1}^d (y_k^{(h)} - \xi_{N,j_k}^{(h)}) (y_\ell^{(\bar{B})} - \xi_{N,j_\ell}^{(\bar{B})}) (y_r^{(\bar{J})} - \xi_{N,j_r}^{(\bar{J})}) \\ &\quad \times \frac{\partial^3 p_1}{\partial y_k^{(h)} \partial y_\ell^{(\bar{B})} \partial y_r^{(\bar{J})}} (\theta y_{-m:u} + (1-\theta) \xi_{N,j-m:u}) , \end{aligned}$$

for a given $\theta \in [0, 1]$ (see [48]). Plugging expansion (45) into (19) leads to:

$$\begin{aligned}
\frac{\bar{p}_{1,N}(\xi_{N,j-m:u})}{p_1(\xi_{N,j-m:u})} &= 1 + \sum_{k=-m}^u \int_{C_{N,j_k}} \frac{\nabla_{y_k} p_1(\xi_{N,j-m:u})^T}{p_1(\xi_{N,j-m:u})} (y_k - \xi_{N,j_k}) \frac{dy_k}{V_{N,j_k}} \\
&+ \frac{1}{2} \sum_{k=-m}^u \int_{C_{N,j_k}} (y_k - \xi_{N,j_k})^T \frac{\nabla_{y_k}^2 p_1(\xi_{N,j-m:u})}{p_1(\xi_{N,j-m:u})} (y_k - \xi_{N,j_k}) \frac{dy_k}{V_{N,j_k}} \\
&+ \frac{1}{2} \sum_{k \neq \ell} \int_{C_{N,j_k}} \int_{C_{N,j_\ell}} (y_k - \xi_{N,j_k})^T \frac{\nabla_{y_k, y_\ell}^2 p_1(\xi_{N,j-m:u})}{p_1(\xi_{N,j-m:u})} (y_\ell - \xi_{N,j_\ell}) \frac{dy_k}{V_{N,j_k}} \frac{dy_\ell}{V_{N,j_\ell}} \\
&+ \epsilon_{N,j-m:u} ,
\end{aligned} \tag{46}$$

where

$$\epsilon_{N,j-m:u} = \int \cdots \int_{C_{N,j-m} \times \cdots \times C_{N,j_u}} \frac{\epsilon_N(y_{-m:u})}{p_1(\xi_{N,j-m:u})} \frac{dy_{-m:u}}{\prod_{i=-m}^u V_{N,j_i}} .$$

We now determine an estimate for this remainder term. For each $y_{-m:u} \in C_{N,j-m} \times \cdots \times C_{N,j_u}$,

$$\begin{aligned}
\frac{\epsilon_N(y_{-m:u})}{p_1(\xi_{N,j-m:u})} &= \frac{1}{6} \sum_{k,\ell,r=-m}^u \sum_{h,\bar{B},\bar{J}=1}^d (y_k^{(h)} - \xi_{N,j_k}^{(h)}) (y_\ell^{(\bar{B})} - \xi_{N,j_\ell}^{(\bar{B})}) (y_r^{(\bar{J})} - \xi_{N,j_r}^{(\bar{J})}) \\
&\times \frac{1}{p_1(\theta y_{-m:u} + (1-\theta)\xi_{N,j-m:u})} \frac{\partial^3 p_1}{\partial y_k^{(h)} \partial y_\ell^{(\bar{B})} \partial y_r^{(\bar{J})}} (\theta y_{-m:u} + (1-\theta)\xi_{N,j-m:u}) \\
&\times \frac{p_1(\theta y_{-m:u} + (1-\theta)\xi_{N,j-m:u})}{p_1(\xi_{N,j-m:u})} . \tag{47}
\end{aligned}$$

First, we find a bound for the last factor. To that end, we expand function $z_{-m:u} \mapsto \log p_1(z_{-m:u})$ at point $\xi_{N,j-m:u}$:

$$\log p_1(z_{-m:u}) = \log p_1(\xi_{N,j-m:u}) + \sum_{k=-m}^u \nabla_{y_k} \log p_1(\theta' z_{-m:u} + (1-\theta')\xi_{N,j-m:u})^T (z_k - \xi_{N,j_k}) ,$$

for a given $\theta' \in [0, 1]$. From Equation (24), the following inequality holds:

$$\begin{aligned}
\left| \log \frac{p_1(z_{-m:u})}{p_1(\xi_{N,j-m:u})} \right| &\leq \sum_{k=-m}^u \left\| \nabla_{y_k} \log p_1(\theta' z_{-m:u} + (1-\theta')\xi_{N,j-m:u}) \right\| \|z_k - \xi_{N,j_k}\| \\
&\leq C_1 \sum_{k=-m}^u \|z_k - \xi_{N,j_k}\| .
\end{aligned}$$

Applying the above upper bound at point $z_{-m:u} = \theta y_{-m:u} + (1-\theta)\xi_{N,j-m:u}$ and using Assumption 3-3), we find

$$\left| \log \frac{p_1(\theta y_{-m:u} + (1-\theta)\xi_{N,j-m:u})}{p_1(\xi_{N,j-m:u})} \right| \leq C_1(m+1) \frac{C_d}{N^{1/d}} ,$$

for each $y_{-m:u} \in C_{N,j-m} \times \cdots \times C_{N,j,u}$. According to the definition of sequence $m(N)$ (see Equation (23)), the r.h.s. of the above equation vanishes as N tends to infinity. Consequently, the term $\frac{p_1(\theta y_{-m:u} + (1-\theta)\xi_{N,j-m:u})}{p_1(\xi_{N,j-m:u})}$ in Equation (47) is bounded. This result together with Assumption 4-2) gives the following upper bound:

$$|\epsilon_{N,j-m:u}| \leq c_T \left(\frac{m+1}{N^{1/d}} \right)^3 ,$$

for some constant c_T .

Let us now examine the dominant terms of Equation (46). Recall that $\xi_{N,j}$ is defined as the centroid of cell $C_{N,j}$:

$$\xi_{N,j} = \int_{C_{N,j}} y \frac{dy}{V_{N,j}} .$$

It is straightforward to prove the following two equalities, for any $j \in \{1, \dots, N\}$ and any d -by- d matrix A :

$$\begin{aligned} \int_{C_{N,j}} (y - \xi_{N,j}) \frac{dy}{V_{N,j}} &= 0 , \\ \int_{C_{N,j}} (y - \xi_{N,j})^T A (y - \xi_{N,j}) \frac{dy}{V_{N,j}} &= \text{Tr}(A M_{N,j}) V_{N,j}^{2/d} . \end{aligned}$$

Plugging above identities in Equation (46) and recalling that $\zeta_{N,j} = \frac{1}{NV_{N,j}}$ prove Lemma 2.

APPENDIX B

PROOF OF LEMMA 3

We study each term of the r.h.s. of (32). Writing Taylor-Lagrange expansions of the probability densities and using the fact that quantization levels are centroids of the cells, we prove the following three lemmas. Define function V_N on \mathcal{Y} by $V_N(y) = V_{N,j}$ whenever $y \in C_{N,j}$.

Lemma 6: For each $k \in \{-m, \dots, u\}$, the following equality holds true:

$$\begin{aligned} \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^T (Y_k - Z_{N,k}) \right] \\ = \frac{1}{N^{2/d}} \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^T \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \frac{\nabla_{y_k} p_0(Y_{-m:k-1}, Z_{N,k}, Y_{k+1:u})}{p_0(Y_{-m:u})} \right] + \bar{\epsilon}_{N,k} , \end{aligned}$$

where $|\bar{\epsilon}_{N,k}| \leq \frac{c'_1}{N^{3/d}}$ for some constant c'_1 .

Proof: We expand the expectation:

$$\begin{aligned} \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^T (Y_k - Z_{N,k}) \right] \\ = \sum_{j_{-m:u}} \int \cdots \int_{C_{N,j_{-m}} \times \cdots \times C_{N,j_u}} \nabla_{y_k} \log p_1(\xi_{N,j_{-m:u}})^T (y_k - \xi_{N,j_k}) p_0(y_{-m:u}) dy_{-m:u} . \end{aligned} \quad (48)$$

where $\sum_{j_{-m:u}}$ is a summation over all index vectors $j_{-m:u} \in \{1, \dots, N\}^{u+m+1}$.

For each $j_k \in \{1, \dots, N\}$, we then consider the Taylor-Lagrange expansion of $y_k \mapsto p_0(y_{-m:u})$ at point ξ_{N,j_k} :

$$\begin{aligned} p_0(y_{-m:u}) &= p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) \\ &\quad + \nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u})^T (y_k - \xi_{N,j_k}) + \epsilon_{N,k}(y_{-m:u}) , \end{aligned} \quad (49)$$

where

$$\epsilon_{N,k}(y_{-m:u}) = (y_k - \xi_{N,j_k})^T \nabla_{y_k}^2 p_0(y_{-m:k-1}, \theta y_k + (1 - \theta) \xi_{N,j_k}, y_{k+1:u}) (y_k - \xi_{N,j_k})$$

for a given $\theta \in [0, 1]$. Under Assumption 4-2), from the counterparts of Equations (24), (25) for density p_0 and following the argument of Lemma 2 (see Appendix A), we can find a bound for this remainder: For each $y_{-m:u} \in C_{N,j_{-m}} \times \cdots \times C_{N,j_u}$,

$$\begin{aligned} |\epsilon_{N,k}(y_{-m:u})| &\leq \|y_k - \xi_{N,j_k}\|^2 \left\| \nabla_{y_k}^2 p_0(y_{-m:k-1}, \theta y_k + (1 - \theta) \xi_{N,j_k}, y_{k+1:u}) \right\| \\ &= \|y_k - \xi_{N,j_k}\|^2 \left\| \frac{\nabla_{y_k}^2 p_0(y_{-m:k-1}, \theta y_k + (1 - \theta) \xi_{N,j_k}, y_{k+1:u})}{p_0(y_{-m:k-1}, \theta y_k + (1 - \theta) \xi_{N,j_k}, y_{k+1:u})} \right\| \\ &\quad \times \frac{p_0(y_{-m:k-1}, \theta y_k + (1 - \theta) \xi_{N,j_k}, y_{k+1:u})}{p_0(y_{-m:u})} p_0(y_{-m:u}) \\ &\leq c \|y_k - \xi_{N,j_k}\|^2 p_0(y_{-m:u}) , \end{aligned} \quad (50)$$

for some constant c .

Plugging expansion (49) into (48) leads to two dominant terms $D_{N,1}$ and $D_{N,2}$ and a remainder r_N :

$$\mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^T (Y_k - Z_{N,k}) \right] = D_{N,1} + D_{N,2} + r_N .$$

We successively study each of them. The first dominant term is

$$\begin{aligned}
D_{N,1} &= \sum_{j-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^T (y_k - \xi_{N,j_k}) \\
&\quad \times p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) dy_{-m:u} \\
&= \sum_{j-m:u} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^T \int \dots \int_{\{C_{N,j_i}\}_{i \neq k}} \left(\int_{C_{N,j_k}} (y_k - \xi_{N,j_k}) dy_k \right) \\
&\quad \times p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) \{dy_i\}_{i \neq k} \\
&= 0,
\end{aligned}$$

where $\{dy_i\}_{i \neq k}$ stands for $\prod_{i=-m, i \neq k}^u dy_i$. The last equality holds true since we have chosen the quantization level $\xi_{N,j}$ to be the centroid of cell $C_{N,j}$.

The second dominant term is

$$\begin{aligned}
D_{N,2} &= \sum_{j-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^T (y_k - \xi_{N,j_k}) (y_k - \xi_{N,j_k})^T \\
&\quad \times \nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) dy_{-m:u} \\
&= \sum_{j-m:u} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^T \int \dots \int_{\{C_{N,j_i}\}_{i \neq k}} \left(\int_{C_{N,j_k}} (y_k - \xi_{N,j_k}) (y_k - \xi_{N,j_k})^T dy_k \right) \\
&\quad \times \nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) \{dy_i\}_{i \neq k} \\
&= \sum_{j-m:u} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^T M_{N,j_k} V_{N,j_k}^{1+2/d} \\
&\quad \times \int \dots \int_{\{C_{N,j_i}\}_{i \neq k}} \nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) \{dy_i\}_{i \neq k}. \quad (51)
\end{aligned}$$

We now write this equality in a simple form. Obviously, under Assumption 1-2), we can write

$$\nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) = \frac{\nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u})}{p_0(y_{-m:u})} p_0(y_{-m:u}).$$

Note that the above expression is independent of $y_k \in C_{N,j}$, so we can also write

$$\nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) = \int_{C_{N,j}} \frac{\nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u})}{p_0(y_{-m:u})} p_0(y_{-m:u}) \frac{dy_k}{V_{N,j}}.$$

Equation (51) thus becomes

$$\begin{aligned}
D_{N,2} &= \sum_{j=-m:u} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^\top M_{N,j_k} V_{N,j_k}^{2/d} \\
&\quad \times \int \dots \int_{\{C_{N,j_i}\}} \frac{\nabla_{y_k} p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u})}{p_0(y_{-m:u})} p_0(y_{-m:u}) dy_{-m:u} \\
&= \frac{1}{N^{2/d}} \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^\top \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \frac{\nabla_{y_k} p_0(Y_{-m:k-1}, Z_{N,k}, Y_{k+1:u})}{p_0(Y_{-m:u})} \right],
\end{aligned}$$

where the last line comes from $\zeta_{N,j} = \frac{1}{NV_{N,j}}$.

We complete the proof with a bound on the remainder term:

$$\begin{aligned}
|r_N| &= \left| \sum_{j=-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} \nabla_{y_k} \log p_1(\xi_{N,j-m:u})^\top (y_k - \xi_{N,j_k}) \epsilon_{N,k}(y_{-m:u}) dy_{-m:u} \right| \\
&\stackrel{(a)}{\leq} C_1 c \int \dots \int \|y_k - \xi_N(y_k)\|^3 p_0(y_{-m:u}) dy_{-m:u} \\
&\stackrel{(b)}{\leq} C_1 c \left(\frac{C_d}{N^{1/d}} \right)^3 = \frac{c'_1}{N^{3/d}},
\end{aligned}$$

where inequality (a) is obtained from Equations (24), (50) and (b) is a consequence of Assumption 3-3).

Putting all pieces together proves Lemma 6. ■

Lemma 7: There exists a constant c'_2 such that, for each $k \neq \ell \in \{-m, \dots, u\}$,

$$\mathbb{E}_0 \left[(Y_k - Z_{N,k})^\top \nabla_{y_k, y_\ell}^2 \log p_1(Z_{N,-m:u}) (Y_\ell - Z_{N,\ell}) \right] \leq \frac{c'_2}{N^{3/d}}.$$

Proof: For each $k \neq \ell$, we expand the expectation:

$$\begin{aligned}
&\mathbb{E}_0 \left[(Y_k - Z_{N,k})^\top \nabla_{y_k, y_\ell}^2 \log p_1(Z_{N,-m:u}) (Y_\ell - Z_{N,\ell}) \right] \\
&= \sum_{j=-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} (y_k - \xi_{N,j_k})^\top \nabla_{y_k, y_\ell}^2 \log p_1(\xi_{N,j-m:u}) (y_\ell - \xi_{N,j_\ell}) p_0(y_{-m:u}) dy_{-m:u}
\end{aligned} \tag{52}$$

and consider the expansion of $y_k \mapsto p_0(y_{-m:u})$ at point ξ_{N,j_k} :

$$p_0(y_{-m:u}) = p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) + \epsilon'_{N,k}(y_{-m:u}), \tag{53}$$

where, from the counterpart of Equation (24) for density p_0 and following the argument leading to Equation (50), $|\epsilon'_{N,k}(y_{-m:u})| \leq c' \|y_k - \xi_{N,j_k}\| p_0(y_{-m:u})$ for some constant c' .

Plugging expansion (53) into (52) leads to a dominant term and a remainder. The dominant term is

$$\begin{aligned}
& \sum_{j=-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} (y_k - \xi_{N,j_k})^\top \nabla_{y_k, y_\ell}^2 \log p_1(\xi_{N,j-m:u}) (y_\ell - \xi_{N,j_\ell}) \\
& \quad \times p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) dy_{-m:u} \\
& = \sum_{j=-m:u} \int \dots \int_{\{C_{N,j_i}\}_{i \neq k}} \left(\int_{C_{N,j_k}} (y_k - \xi_{N,j_k}) dy_k \right) \nabla_{y_k, y_\ell}^2 \log p_1(\xi_{N,j-m:u}) (y_\ell - \xi_{N,j_\ell}) \\
& \quad \times p_0(y_{-m:k-1}, \xi_{N,j_k}, y_{k+1:u}) \{dy_i\}_{i \neq k} \\
& = 0 .
\end{aligned}$$

Using Equation (25) and Assumption 3-3), we find a bound for the remainder term:

$$\begin{aligned}
& \left| \sum_{j=-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} (y_k - \xi_{N,j_k})^\top \nabla_{y_k, y_\ell}^2 \log p_1(\xi_{N,j-m:u}) (y_\ell - \xi_{N,j_\ell}) \epsilon'_{N,k}(y_{-m:u}) dy_{-m:u} \right| \\
& \leq C_2 c' \left(\frac{C_d}{N^{1/d}} \right)^3 = \frac{c'_2}{N^{3/d}} . \quad (54)
\end{aligned}$$

■

Lemma 8: For each $k \in \{-m, \dots, u\}$,

$$\begin{aligned}
& \mathbb{E}_0 \left[(Y_k - Z_{N,k})^\top \nabla_{y_k}^2 \log p_1(Z_{N,-m:u}) (Y_k - Z_{N,k}) \right] \\
& = \frac{1}{N^{2/d}} \mathbb{E}_0 \left[\text{Tr} \left(\nabla_{y_k}^2 \log p_1(Z_{N,-m:u}) \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \right) \frac{p_0(Y_{-m:k-1}, Z_{N,k}, Y_{k+1:u})}{p_0(Y_{-m:u})} \right] + \epsilon'_{N,k} ,
\end{aligned}$$

where $|\epsilon'_{N,k}| \leq \frac{c'_2}{N^{3/d}}$.

Proof: For each k , we expand the expectation:

$$\begin{aligned}
& \mathbb{E}_0 \left[(Y_k - Z_{N,k})^\top \nabla_{y_k}^2 \log p_1(Z_{N,-m:u}) (Y_k - Z_{N,k}) \right] \\
& = \sum_{j=-m:u} \int \dots \int_{C_{N,j-m} \times \dots \times C_{N,j_u}} (y_k - \xi_{N,j_k})^\top \nabla_{y_k}^2 \log p_1(\xi_{N,j-m:u}) (y_k - \xi_{N,j_k}) p_0(y_{-m:u}) dy_{-m:u} .
\end{aligned} \quad (55)$$

Plugging expansion (53) into (55) leads to a dominant term and a remainder. The study of the dominant term uses the same arguments as Lemma 6. The final expression comes from the

following equality:

$$\int_{C_{N,j}} (y - \xi_{N,j})^T A (y - \xi_{N,j}) dy = \text{Tr}(A M_{N,j}) V_{N,j}^{1+2/d},$$

for any d -by- d matrix A , and the definition of the specific point density $\zeta_{N,j} = \frac{1}{NV_{N,j}}$.

Equation (54) is also valid when $k = \ell$ *i.e.*, for the remainder considered here. This proves Lemma 8. ■

Gathering Equation (32) and Lemmas 6, 7, 8 results in

$$\begin{aligned} U_N(u) = & -\frac{1}{N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\nabla_{y_k} \log p_1(Z_{N,-m:u})^T \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \frac{\nabla_{y_k} p_0(Y_{-m:k-1}, Z_{N,k}, Y_{k+1:u})}{p_0(Y_{-m:u})} \right] \\ & - \frac{1}{2N^{2/d}} \sum_{k=-m}^u \mathbb{E}_0 \left[\text{Tr} \left(\nabla_{y_k}^2 \log p_1(Z_{N,-m:u}) \frac{M_N(Y_k)}{\zeta_N(Y_k)^{2/d}} \right) \frac{p_0(Y_{-m:k-1}, Z_{N,k}, Y_{k+1:u})}{p_0(Y_{-m:u})} \right] \\ & + \bar{\epsilon}_N(u), \end{aligned}$$

where $|\bar{\epsilon}_N(u)| \leq c_U \frac{m^3}{N^{3/d}}$ for some constant c_U .

Expanding $\nabla_{y_k} p_0$ and p_0 once again, under Assumptions 3 and 4-2), it is straightforward to write the dominant term in a simple form *i.e.*, replace each $Z_{N,k}$ by Y_k . From Equation (23), the remainder term is a little-o of $N^{-2/d}$. This proves Lemma 3.

APPENDIX C

PROOF OF LEMMA 5

Equation (38) ensures that the following series converges:

$$\Sigma_N = \mathbb{E}_0 [\mathcal{H}_{N,0}(Y_{-\infty:0})] + \sum_{k=-\infty}^{-1} \mathbb{E}_0 [\mathcal{H}_{N,k}(Y_{-\infty:0}) - \mathcal{H}_{N,k}(Y_{-\infty:-1})].$$

Using Equation (35), the approximation of $N^{2/d}(K - K_N)$ by series Σ_N leads to the following remainder:

$$|N^{2/d}(K - K_N) - \Sigma_N| \leq \sum_{k=-m}^0 \mathbb{E}_0 |\Delta_N^{(k)}| + \sum_{k=-\infty}^{-m-1} \mathbb{E}_0 |\Upsilon_N^{(k)}| + \epsilon_N, \quad (56)$$

where $\Delta_N^{(0)} = \mathcal{H}_{N,0}(Y_{-m:0}) - \mathcal{H}_{N,0}(Y_{-\infty:0})$ and

$$\Delta_N^{(k)} = \mathcal{H}_{N,k}(Y_{-m:0}) - \mathcal{H}_{N,k}(Y_{-m:-1}) - \mathcal{H}_{N,k}(Y_{-\infty:0}) + \mathcal{H}_{N,k}(Y_{-\infty:-1}) \quad (\forall k \leq -1),$$

$$\Upsilon_N^{(k)} = \mathcal{H}_{N,k}(Y_{-\infty:0}) - \mathcal{H}_{N,k}(Y_{-\infty:-1}) \quad (\forall k \leq -m-1),$$

and where $\check{\epsilon}_N \rightarrow 0$ as $N \rightarrow \infty$. Using the triangular inequality, we obtain for each $k \leq -1$:

$$\mathbb{E}_0 |\Delta_N^{(k)}| \leq \mathbb{E}_0 |\mathcal{H}_{N,k}(Y_{-m:0}) - \mathcal{H}_{N,k}(Y_{-\infty:0})| + \mathbb{E}_0 |\mathcal{H}_{N,k}(Y_{-m:-1}) - \mathcal{H}_{N,k}(Y_{-\infty:-1})| .$$

Using (37), this leads to:

$$\mathbb{E}_0 |\Delta_N^{(k)}| \leq 2 c_h \varphi_{m-|k|} .$$

From the triangular inequality once again,

$$\mathbb{E}_0 |\Delta_N^{(k)}| \leq \mathbb{E}_0 |\mathcal{H}_{N,k}(Y_{-m:0}) - \mathcal{H}_{N,k}(Y_{-m:-1})| + \mathbb{E}_0 |\mathcal{H}_{N,k}(Y_{-\infty:0}) - \mathcal{H}_{N,k}(Y_{-\infty:-1})| .$$

Using (38), this leads to:

$$\mathbb{E}_0 |\Delta_N^{(k)}| \leq 2 c_h \psi_{|k|} .$$

After some algebra, there exists a constant c_Δ such that:

$$\begin{aligned} \sum_{k=-m}^{-1} \mathbb{E}_0 |\Delta_N^{(k)}| &\leq c_\Delta \sum_{k=-m}^{-1} \varphi_{m-|k|} \wedge \psi_{|k|} \\ &\leq c_\Delta \left(\sum_{k=-\lfloor m/2 \rfloor}^{-1} \varphi_{m-|k|} + \sum_{k=-m}^{-\lfloor m/2 \rfloor} \psi_{|k|} \right) \\ &\leq c_\Delta \left(\sum_{k=\lceil m/2 \rceil}^{\infty} \varphi_k + \sum_{k=\lfloor m/2 \rfloor}^{\infty} \psi_k \right) \\ &\leq c_\Delta \mathcal{T}_\Delta^{(m)} , \end{aligned}$$

Where $(\mathcal{T}_\Delta^{(m)})_{m \geq 0}$ is a sequence of positive numbers such that $\mathcal{T}_\Delta^{(m)} \rightarrow 0$ as $m \rightarrow \infty$. The last line of the above equation holds true under Assumption 4-4) since $\sum \varphi_k$ and $\sum \psi_k$ are convergent series. Similarly, $\mathbb{E}_0 |\Delta_N^{(0)}| \leq c_h \varphi_m$.

The last series in (56) can be bounded using (38):

$$\sum_{k=-\infty}^{-m-1} \mathbb{E}_0 |\Upsilon_N^{(k)}| \leq c_h \sum_{k=-\infty}^{-m-1} \psi_{|k|} = c_\Upsilon \mathcal{T}_\Upsilon^{(m)} ,$$

for some constant c_Υ and a given sequence $(\mathcal{T}_\Upsilon^{(m)})_{m \geq 0}$ such that $\mathcal{T}_\Upsilon^{(m)} \rightarrow 0$ as $m \rightarrow \infty$.

Putting all pieces together, Equation (56) leads to:

$$|N^{2/d}(K - K_N) - \Sigma_N| \leq c_h \varphi_m + c_\Delta \mathcal{T}_\Delta^{(m)} + c_\Upsilon \mathcal{T}_\Upsilon^{(m)} + \check{\epsilon}_N .$$

The r.h.s. of the above inequality tends to zero as $m, N \rightarrow \infty$. This proves Lemma 5.

ACKNOWLEDGMENT

The authors would like to thank Prof. Eric Moulines for helpful comments and for bringing useful references to their attention. They are also grateful to Dr. Walid Hachem and Dr. Pablo Piantanida for fruitful discussions.

REFERENCES

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [2] B. Chen, L. Tong, and P. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 1053, no. 5888/06, pp. 16–26, 2006.
- [3] R. Gray and D. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [4] A. Gersho and R. Gray, *Vector quantization and signal compression*. Kluwer, 1992.
- [5] W. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, vol. 27, pp. 446–472, 1948.
- [6] S. Na and D. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 886–900, 1995.
- [7] T. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, 1998.
- [8] V. Misra, V. Goyal, and L. Varshney, "Distributed functional scalar quantization: High-resolution analysis and extensions," *Arxiv*, vol. cs.IT, p. arXiv:0811.3617, 2008.
- [9] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, 2006.
- [10] K. Perlmutter, S. Perlmutter, R. Gray, R. Olshen, and K. Oehler, "Bayes risk weighted vector quantization with posterior estimation for image compression and classification," *IEEE Trans. Image Process.*, vol. 5, no. 2, pp. 347–360, 1996.
- [11] S. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. 25, no. 5, pp. 479–484, 1977.
- [12] H. Poor and J. Thomas, "Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. 25, no. 9, pp. 893–900, 1977.
- [13] H. Poor, "Fine quantization in signal detection and estimation," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 960–972, 1988.
- [14] B. Picinbono and P. Duvaut, "Optimum quantization for detection," *IEEE Trans. Commun.*, vol. 36, no. 11, pp. 1254–1258, 1988.
- [15] J. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 550–558, 1993.
- [16] R. Tenney and N. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, no. 4, pp. 501–510, 1981.
- [17] J. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals, and Systems*, vol. 1, no. 2, pp. 167–182, 1988.
- [18] R. Gupta and A. Hero, "High-rate vector quantization for detection," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1951–1969, 2003.
- [19] E. Lehmann and J. Romano, *Testing Statistical Hypotheses (3rd Ed)*. Springer Texts in Statistics, 2005.
- [20] T. Cover and J. Thomas, *Elements of information theory (2nd Ed)*. Wiley-Interscience, 2006.

- [21] J.-F. Chamberland and V. Veeravalli, “How dense should a sensor network be for detection with correlated observations?” *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, 2006.
- [22] P. Willett, P. Swaszek, and R. Blum, “The good, bad, and ugly: Distributed detection of a known signal in dependent Gaussian noise,” *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3266–3279, 2000.
- [23] Y. Sung, L. Tong, and H. Poor, “Neyman–Pearson detection of Gauss–Markov signals in noise : closed-form error exponent and properties,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1354–1365, 2006.
- [24] W. Hachem, E. Moulines, and F. Roueff, “Error exponents for Neyman–Pearson detection of a continuous-time Gaussian Markov process from noisy irregular samples,” *arXiv cs.IT*, 2009, submitted to IEEE Trans. on Inf. Theory.
- [25] P.-N. Chen, “General formulas for the Neyman–Pearson type-II error exponent subject to fixed and exponential type-I error bounds,” *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 316–323, 1996.
- [26] R. Bradley, “Basic properties of strong mixing conditions. a survey and some open questions,” *Probability Surveys*, vol. 2, pp. 107–144, 2005.
- [27] S. Moy, “Generalizations of Shannon–McMillan theorem,” *Pacific J. Math.*, vol. 11, no. 2, pp. 705–714, 1961.
- [28] P. Doukhan, *Mixing: properties and examples*. Springer, 1994.
- [29] D. Bosq, *Nonparametric statistics for stochastic processes: estimation and prediction*. Springer Verlag, 1998.
- [30] J. Villard and P. Bianchi, “High-rate vector quantization for the Neyman–Pearson detection of some stationary mixing processes,” in *ISIT*, Austin, Texas, USA, 2010.
- [31] P. Panter and W. Dite, “Quantization distortion in pulse-count modulation with nonuniform spacing of levels,” *Proceedings of the IRE*, vol. 39, no. 1, pp. 44 – 48, 1951.
- [32] D. Neuhoff, “On the asymptotic distribution of the errors in vector quantization,” *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 461–468, 1996.
- [33] A. Gersho, “Asymptotically optimal block quantization,” *IEEE Trans. Inf. Theory*, vol. 25, no. 4, pp. 373–380, 1979.
- [34] R. Zamir and M. Feder, “On lattice quantization noise,” *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1152 –1159, 1996.
- [35] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, 1980.
- [36] J. Conway and N. Sloane, *Sphere packings, lattices, and groups (3rd Ed)*. Springer-Verlag, 1999.
- [37] R. Gupta, “Quantization strategies for low-power communications,” Ph.D. dissertation, The University of Michigan, 2001.
- [38] J. Lasserre, “A trace inequality for matrix product,” *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1500 –1501, 1995.
- [39] A. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications*. Academic Press New York, 1979.
- [40] P. Billingsley, *Probability and Measure (3rd Ed)*. John Wiley & Sons, 1995.
- [41] R. Douc, E. Moulines, and T. Ryden, “Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime,” *The Annals of Statistics*, vol. 32, no. 5, pp. 2254–2304, 2004.
- [42] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*. Springer series in statistics, 2007.
- [43] J. Proakis and M. Salehi, *Digital communications (5th Ed)*. McGraw-Hill, 2007.
- [44] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions, vol. 1 (2nd Ed)*. Wiley-Interscience, 1994.
- [45] J. Liu, *Monte Carlo strategies in scientific computing*. Springer Verlag, 2001.
- [46] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*. Springer, 2000.
- [47] A. Dembo and O. Zeitouni, *Large deviations techniques and applications (2nd Ed)*. Springer Verlag, 1998.
- [48] S. Lang, *Calculus of several variables*. Addison-Wesley, 1973.

Joffrey Villard (S'09) was born in Saint-Étienne, France, in 1985. He received the Dipl.Ing. degree in digital communication and electronics, and the M.Sc. degree in wireless communication systems, both from Supélec, Gif-sur-Yvette, France, in 2008.

He is currently working towards the Ph.D. degree at the Department of Telecommunications of Supélec. His research interests include information theory, source coding, statistical inference, and signal processing for wireless sensor networks.

Pascal Bianchi (M'06) was born in 1977 in Nancy, France. He received the M.Sc. degree of Supélec-Paris XI in 2000 and the Ph.D. degree of the University of Marne-la-Vallée in 2003.

From 2003 to 2009, he was an Associate Professor at the Telecommunication Department of Supélec. In 2009, he joined the Statistics and Applications group at LTCI-Telecom ParisTech. His current research interests are in the area of statistical signal processing for sensor networks. They include decentralized detection, quantization, stochastic optimization, and applications of random matrix theory.